utterance to a text message. The command or request may be compared against a context description grammar to identify a match. Any active grammars in the context description grammar may be scored against the command or request and a best match may be sent to a response generator module. Agents may be associated with corresponding response generator modules and may retrieve the requested information for generation of a response. The agents may update a context stack to enable follow-up requests.

[021] According to another embodiment of the invention, mobile devices may be configured to allow verbal annotations of objects stored thereon. Mobile devices may transcribe the verbal annotation to text and store the textual annotation with the object. Alternatively, mobile devices may be configured to enable users to manually input textual descriptions that are stored along with the objects. According to one embodiment of the invention, the textual annotations and/or textual descriptions may be classified and searched. In an alternative embodiment, mobile devices may classify and search the verbal annotations rather than textual annotations. However, classifying and searching verbal annotations may be considerably more difficult than classifying and searching textual annotations and/or textual descriptions.

[022] According to one embodiment of the invention, the textual annotations and textual descriptions may be communicated using a short message service on the mobile telephone or other device. Short message service is a text message service that enables sending and receiving of short textual messages. The textual messages may be stored at data centers for forwarding to intended recipients. Other configurations may be used.

[023] According to another aspect of the invention, domain specific behavior and information may be organized into data managers. Data managers are autonomous executables that receive, process and respond to user questions, queries and commands. The data managers provide complete, convenient and re-distributable packages or modules of functionality that are typically directed to a specific domain of applications. Data managers may be complete packages of executable code, scripts, links to information, and other forms of communication data that provide a specific package of functionality, usually in a specific domain. In other words, data managers may include components for extending the functionality to a new domain. Further, data managers and their associated data may be updated remotely over a

Google Exhibit 1019
Page 00501

network as new behavior is added or new information becomes available. Data managers may use system resources and the services of other, typically more specialized, data managers. Data managers may be distributed and redistributed in a number of ways including on removable storage media, transfer over networks or attached to emails and other messages. An update manager may be used to add new data managers to the system or update existing data managers.

[024] In order to enhance the natural query and response environment, the system may format results to increase understandability to users. Formatting and presentation of results may be based on the context of the questions, the contents of the response being presented, the history of the interaction with the user, the user's preferences and interests and the nature of the domain. By contrast, rigid, highly formatted, or structured presentation of results may be deemed unnatural by many users.

[025] According to another embodiment of the invention, the system may simulate some aspects of a human "personality". In some cases, the presentation of the response and the terms that are used to provide the response may be randomized to avoid the appearance of rigidly formatted or mechanical responses. The use of other simulated personality characteristics is also desirable. For example, a response that may be upsetting to the user may be presented in a sympathetic manner. Furthermore, results of requests may be long text strings, lists, tables or other lengthy sets of data. Natural presentation of this type of information presents particular challenges because simply reading the long response is generally not preferred. Instead, the system may parse important sections from the response and may initially provide only reports. Determining what parts of a long response are presented may be based on the context of the questions, the contents of the response being presented, the history of the interaction with the user, the user's preferences and interests and the nature of the domain. At the same time, the system may give the user interactive control over what information to present and how much information to present, to stop the response all together, or to take other actions.

[026] The invention may be applied as a user interface to telematics systems in a wide variety of environments. These environments may include, but are not limited to, the following: 1) personal automobiles, rented automobiles, or fleet automobiles; 2) motorcycles, scooters, and other two wheeled or open-air vehicles; 3)commercial long-haul and short haul trucks; 4) delivery service vehicles; 5) fleet service vehicles;

Google Exhibit 1019
Page 00502

6) industrial vehicles; 7) agricultural and construction machinery; 8) water-borne vehicles; 9) aircraft; and, 10) specialized military, law enforcement and emergency vehicles.

[027] According to another aspect of the invention, the system may process and respond to questions, requests and/or commands. Keywords or context may be used to determine whether the received utterance and/or textual message includes a request or command. For example, utterances may include aspects of questions, requests and/or commands. For example, a user may utter "tune in my favorite radio station." A request is processed to determine the name, the channel, and time for the users favorite radio station. If the programming on that station is of a type the user generally does not listen to, the system can suggest using an alterative option, such as listening to a CD more likely to please the user. A command must be executed to set a tune the radio station.

[028] The invention can be used for generalized local or network information query, retrieval and presentation in a mobile environment. For each user utterance, including a question or query or set of questions or queries, the system may perform multiple steps possibly including: 1) capturing the user's question or query through speech recognition operating in a variety of real-world environments; 2) parsing and interpreting the question or query; 3) determining the domain of expertise required and the context to invoke the proper resources, including agents; 4) formulating one or more queries to one or more local and/or network data sources or sending appropriate commands to local or remote devices or the system itself; 5) performing required formatting, variable substitutions and transformations to modify the queries to a form most likely to yield desired results from the available sources; 6) executing the multiple queries or commands in an asynchronous manner and dealing gracefully with failures; 7) extracting or scraping the desired information from the one or more results, which may be returned in any one of a number of different formats; 8) evaluating and interpreting the results, including processing of errors, gathered and combine them into a single best result judged to be "best" even if the results are ambiguous, incomplete, or conflicting; 9) performing required formatting, variable substitutions and transformations to modify the results to a form most easily understood by the user; 10) presenting the compound result, through a text to speech engine or multimodal interface, to the user in a useful and/or expected manner; 11)

Google Exhibit 1019
Page 00503

optionally, providing a response to users indicating the success or failure of the command, and possibly including state information; or other steps.

[029] The above steps may be performed with knowledge of the domain of expertise, the context for the question or command, domain specific information, the history of the user's interactions, user preferences, available information sources or commands, and responses obtained from the sources.

[030] Probabilistic or fuzzy set decision and matching methods may be applied to deal with inconsistent, ambiguous, conflicting and incomplete information or responses. In addition, asynchronous queries may be used to provide rapid and graceful failure of requests or commands that allow the system to robustly return results quickly and in a manner that seems natural to the user.

[031] Many everyday questions are inherently subjective and result in answers that are a matter of opinion or consensus, as much as fact. Such questions are often ad hoc in their nature, as well. The system, according to another aspect of the invention, may use adaptive, probabilistic, and fuzzy set decision and matching methods to identify the subjective nature of the question and to evaluate a range of possible answers, wherein one or more answers may be selected that most accurately represent the type of result desired by the user.

[032] The context and expected results from a particular question may be highly dependent on the individual asking the question. Therefore, the system may create, store and use personal profile information for each user. Information in the profile may be added and updated automatically as the user uses the system or may be manually added or updated by the user or by others. Domain specific agents may collect, store and use specific profile information, as may be required for optimal operations. Users may create commands for regularly used reports, automatically generated alerts, and other requests and for the formatting and presentation of results. The system may use profile data in interpreting questions, formulating requests, interpreting request results and presenting answers to the user. Examples of information in a user profile includes, history of questions asked, session histories, formatting and presentation preferences, special word spelling, terms of interest, special data sources of interest, age, sex, education, location or address, place of business, type of business, investments, hobbies, sports interests, news interests and other profile data.

Google Exhibit 1019
Page 00504

[033] According to one aspect of the invention, the system may attempt to provide rapid responses in order to provide a natural question and response environment. The rapid responses may be provided without obtaining additional information. The system may determine agent composition, context and/or domain for a user's question or command, for example, by using a real-time scoring system or other technique. Based on this determination, the system may trigger one or more agents to respond to the user's question or command. The agents may make one or more requests and rapidly return a formatted response. Thus, users may receive direct responses to a set of questions, each with a different response or context. In some cases, the available information, including the request results, may not adequately answer the questions presented. In such situations, the user may be asked one or more follow-up questions to resolve the ambiguity. Additional requests may then be made before an adequate response is provided. In these cases, the system may use context information, user profile information and/or domain specific information to minimize the interaction with the user required to deliver a response.

[034] If the confidence level of the domain or context score is not high enough to ensure a reliable response, the system may request that the user verify that the question or command is correctly understood. In general, the question may be phrased to indicate the context of the question including all criteria or parameters. If the user confirms that the question is correct, the system may proceed to produce a response. Otherwise, either the user can rephrase the original question, perhaps adding additional information to remove ambiguity, or the system may ask one or more questions to attempt to resolve the ambiguity or other actions may taken.

[035] According to one aspect of the invention, the system may accept any natural language question or command and, as a result, may be subject to ambiguous requests. To assist users in formulating concise questions and commands, the system may support a voice query language. The language may help users clearly specify the keywords or contexts of the question or command along with the parameters or criteria. The system may provide built in training capabilities to help the user learn the best methods to formulate their questions and commands.

[036] In order to make the responses to user's questions and commands seem more natural, the system may employ one or more dynamically invokeable personalities and/or emotional models. Personalities and emotional models have specific

Google Exhibit 1019
Page 00505

characteristics that simulate the behavioral characteristics of real humans. Examples of these characteristics include, sympathy, irritation, and helpfulness and associated emotions. The personality also randomizes aspects of responses, just as a real human would do. This behavior includes randomization of terms used and the order of presentation of information. Characteristics of the personality and/or emotions are invoked using probabilistic or fuzzy set decision and matching methods and using criteria, including the context for the question, the history of the user's interaction, user preferences, information sources available, responses obtained from the sources.

[037] According to another aspect of the invention, special procedures may be employed to present information in the form of long text strings, tables, lists or other long response sets. Simply presenting a long set of information in an ordered manner may not be considered natural or what most users have in mind. The system may use probabilistic or fuzzy set matching methods to extract relevant information and present these subsets first. Further the system may provide commands allowing users to skip through the list, find keywords or key information in the list or stop processing the list altogether.

[038] According to one embodiment of the invention, the system may support multiple users that access the system at different times. According to another embodiment of the invention, the system may support multiple users that access the system during a same session, in an interleaved or overlapping manner. The system may recognize the multiple users by name, voice, or other characteristic and may invoke a correct profile for each user. If multiple users are addressing the system in overlapping or interleaved sessions, the system may identify the multiple users and may invoke one or more corresponding profiles. For applications that require security safeguards, the multiple users may be verified using voiceprint matching, password or pass-phrase matching, or other security safeguards.

[039] When multiple users are engaged in interleaved sessions, the system may gracefully resolve conflicts using a probabilistic or fuzzy set decision method for each user. This process may simulate the manner in which a human would address multiple questions from various sources. For example, the system may answer short questions first in time while answering longer questions later in time. Alternatively, the system may answer questions in the order that they are received, among other configurations.

Google Exhibit 1019
Page 00506

[040] Since the invention may operate in many environments, including mobile environments with background noise, point noise sources and people holding conversations, filtering of speech input may be advantageous. The invention may use, for example, either one-dimensional or two-dimensional array microphones (or other devices) to receive human speech. The array microphones may be fixed or employ dynamic beam forming techniques. The array pattern may be adjusted to maximize gain in the direction of the user and to null point noise sources. Alternatively, microphones may be placed at particular locations within the mobile environment near where occupants are likely to use the system. These microphones can be single microphones, directional microphones or an array of microphones. Speech received at the microphones may then be processed with analog or digital filters to optimize the bandwidth, cancel echoes, notch-out narrow band noise sources, or perform other functions. Following filtering, the system may use variable rate sampling to maximize the fidelity of the encoded speech, while minimizing required bandwidth. This procedure may be particularly useful in cases where the encoded speech is transmitted over a wireless network or link.

[041] The invention can be applied to a wide range of telematics applications. General applications areas can include, but are not limited to remote or local vehicle control, information query, retrieval and presentation from local or network sources, safety applications, and security applications.

[042] The system may provide local or remote control functions for the system, including devices that are located on the mobile structure or off the mobile structure. Users may initiate commands locally or remotely. Typically, remote operation may be conducted through an IP connection, a telephone connection, or other connections. The user may address spoken commands to a mobile device or desktop unit, which may send the commands to controllers on the vehicle over wireless links. Other remote command techniques may be used. The system may process commands in a nearly identical manner to a request. One difference being that the result of the command is generally an action rather than a response. In many cases, the system may give the user a cue or response to indicate that the command has been successfully executed or has failed. In cases of failure, an interactive session may be started to allow the user to resolve the difficulty or formulate a command more likely to succeed.

Google Exhibit 1019
Page 00507

[043] The invention provides users, including mobile structure operators, with the capability to control most any mobile system functions using interactive speech and non-speech commands and/or requests. Generally, controls of a critical nature, or having safety implications, may employ fail-safe checks that, before execution, verify that a command will not create a hazardous condition. Manual overrides may also be provided as an extra precaution. The invention may provide built in help and user guidance for the devices under control. This guidance may include step-by-step training for operators that are learning to use the features of the mobile structure. The system can provide extensive interactive guidance when commands cannot be executed or when commands fail. This advice may include suggestions to reformulate the command to succeed, suggestions to work around a failure, suggestions for alternative commands that may achieve a similar function, or other suggestions. Examples of control functions that can be performed from local or remote locations by the invention include:

1. Control of mobile structure multimedia entertainment electronics, such as radio, CD player, or video player. This control can be based on user specified play lists and may be sensitive to user profile information including, preferences history, or other information. The invention may include the ability to control multiple or individual multimedia entertainment stations.
2. Control of communications devices such as cell phones, voice mail systems, fax systems, text or instant messaging systems, call and message forwarding systems, email systems and other communication devices. This control includes controlling features such as, address books, phone books, call forwarding, conference calling, and voice mail, among others.
3. Local or remote control of mobile structure systems. Most any device on the mobile structure may be controlled including, door locks, window controls, interior temperature controls, shifting of the transmission, turn signals, lights, safety equipment, engine ignition, cruse control, fuel tank switches, seat adjustments, specialized equipment such as winches, lifting systems or loading systems, and other mobile structure systems.
4. Systems that are located external to the mobile structure typically may be controlled through wireless links including, garage door openers, gate

Google Exhibit 1019
Page 00508

controllers, mobile structure entry security passes, automated toll collection systems, and mobile structure weighing systems and other external systems.

5. Mobile structure power management and systems control. The invention may provide the mobile structure operator with information on limitations and on tips for better power management or fuel utilization or other systems control.

6. Diagnostic information management. The invention may provide diagnostic information announcements and warnings for the mobile structure operator. These announcements and warnings may be interactive allowing the operator to request additional information or to suggest various courses of action. The invention may mediate a solution to the problem, including providing access to a scheduling service, summoning help or providing instructions for remedial action until a permanent solution can be achieved. The system can ask the operator to authorize the ordering of needed parts and may provide cost estimates. The system may receive data for these announcements and warnings from a wide range of sources including sensors and vehicle control computers. Sensors may include fuel level sensors, coolant temperature sensors, oil temperature sensors, axel temperature sensors, tire pressure sensors, and other sensors.

7. System status inquiry. Mobile structure operators may use the interactive natural language interface of the invention to query and receive reports on the status of systems that are located on the mobile structure including, fuel level, interior temperature, outside temperature, engine or other mobile structure system status. The operator can further query the system to receive more information or determine a course of action if problem is detected.

8. Vehicle service history. The invention can provide the vehicle operator or other personnel with interactive access to the vehicle service history. The invention can provide announcements or warnings as the time for service draws near. The user can interact with the system to schedule the required service, order required parts, receive cost estimates, or update the service history. Users may customize the nature of this interaction to suit their desires or policies.

Google Exhibit 1019
Page 00509

9. Diagnostic and service history. The invention may provide diagnostic and service history information to service personnel. This information may include vehicle fault codes and other information for devices that are under control of or measured by the system. Alternatively, the invention may receive information on the state and history of vehicle operation from other control computers. The invention may provide interactive service information and history. The service history may be queried and presented through the speech interface or the non-speech interface. The system may prompt service personnel for more information if the record is deemed incomplete. In other cases, the invention may prompt service personnel for information on their actions if a change is detected in system status, such as replacement of a part.

[044] The invention can provide users or operators of a mobile structure with specialized safety functions through the interactive speech interface and/or non-speech interface. The invention may use a dynamically evocable personality that is capable of creating announcements that are appropriate for the severity of the situation. The announcements and personalities may be under user control and configuration. Some examples of these safety applications can include:

1. The invention can provide automated detection and reporting of accident situations through the wireless communications link. Information on an accident situation can be gathered from airbag control systems or other sensors. Once an accident situation has been detected the invention may use the interactive speech interface and/or non-speech interface to determine the nature of the accident and/or the condition of victims. This information, along with location information and other relevant information, may be reported over the wireless link. Alternatively, the invention may establish voice channel communications or other communications between occupants of the vehicle and emergency personnel.

2. The invention can be used to store and retrieve medical information on vehicle occupants. Following an accident, emergency personnel may query the system for this information. Alternatively, the system may annunciate a warning to emergency personnel if a person has a special medical condition. The system maintains the security of medical information through a number of techniques, including not annunciating medical information unless an accident

Google Exhibit 1019
Page 00510

is detected, or not annunciating medical information unless that person or another authorized person gives permission.

3. Occupants of the vehicle can summon help in the event of a crime using the speech interface and/or non-speech interface. Typical crimes may include robberies and hijackings. The invention allows vehicle occupants to set panic or emergency words or phrases that indicate to the system a crime is occurring.

4. The invention can provide the vehicle operator with safety announcements if an unsafe or potentially unsafe situation is detected. The operator can use the interactive speech interface and/or non-speech interface to obtain more information on the situation or dismiss the alert. The operator can annunciate or otherwise provide commands to remedy or mitigate the situation during this dialog. Conditions that can be announced include, following another vehicle too closely, too great a speed for the road or conditions, obstruction on roadway, a fire in some part of the vehicle, high cargo pressure or temperature, leaks, and other information.

5. The interactive speech interface and non-speech interface of the invention can provide the operator with real-time assistance. This assistance may include, aid in parking or backing a vehicle, aid with complex maneuvers, aid with optimal operation of the vehicle, and other operations. The operator can ask the system for advice or assistance with a planned maneuver or operation. Alternatively, the invention can proactively offer assistance if certain situations are detected.

6. The interactive speech interface and non-speech interface may be used to improve vehicle security. Voiceprints or voice authentication can be used to gain access to the vehicle or start the vehicle. Alternatively or in addition, a password or pass-phrase can be used. In another alternative speech security can be used as a supplement to other vehicle security techniques.

7. The invention can provide measurement of operator fatigue and alert the operator or remote personnel if unacceptable levels of fatigue are detected. The interactive speech interface and/or non-speech interface can be used to query the operator to detect fatigue. Alternatively, or in addition, other measurements of operator fatigue can be used. If a fatigue situation is

Google Exhibit 1019
Page 00511

detected the invention may initiate a dialog with the operator to determine the extent of the problem and if required, ask the operator to cease operation.

[045] The invention can offer vehicle operators and occupants a variety of services that are useful while in the vehicle and/or while arriving at a destination. Further, users can employ the interactive natural language interface to customize these services to suit each individual. Some examples of services that can be supported by the natural-language interactive speech interface of the invention, include:

1. Providing vehicle operators with interactive directions to a destination or waypoint, wherein the user can specify a desired destination and any preferred waypoints. A destination can be specified in any manner including providing the name of a place, an address, a name of a person, a name of a business, or other type of information. As the trip progresses the system may provide the operator with continued directions and warnings if a mistake has been made. The operator can query the system for additional information or request less information. Generally, the system is interfaced with one or more navigation sensors and local or remote map databases. The invention may provide operators or passengers with alerts of upcoming points or interest, exits or stops, hazards, or other conditions. The users can query the system for more specific information. Alternatively, the invention can provide operators and occupants of the vehicle with an interactive guided tour. The system's information query, retrieval and presentation capability can be employed by users to receive additional information or points or items of interest during the tour and may take into account stored personal profile information for a user.
2. The invention can provide the operator of a vehicle with interactive dynamic routing information. The routing can be updated based on traffic conditions, weather conditions, facilities availability, and information provided by the operator. Generally, the system is interfaced with one or more navigation sensors, local or remote map databases, and sources of traffic, weather, and facilities use data.
3. The direction, routing and communications capabilities of the invention can be combined in an interactive system that helps one or more operators rendezvous at a predetermined destination or any other convenient midpoint. The operators may use the interactive natural language interface to

Google Exhibit 1019
Page 00512

communicate with the system to arrange the rendezvous, to receive directions as they travel to the rendezvous point and/or to communicate with the other operators.

4. The navigational capabilities of the invention may be used to place limits on where a vehicle is allowed to go and/or for how long. The system may employ the interactive natural language speech interface and/or non-speech interface to inform the operator when the vehicle is approaching or has exceeded a limit. The operator can query the system to determine the best course of action to return to limits or prevent exceeding them. Alternatively, the system can query the operator to determine why they are exceeding the limits or to mediate a negotiation to extend the limits if this is required by circumstances. This capability is useful in several situations including, keeping a delivery or passenger vehicle on a regular route, setting and enforcing use limits on teenagers, preventing an operator from using the vehicle in an unauthorized manner, or other situations.

5. The interactive natural language interface of the invention can be used to provide Customer Relationship Management (CRM) services to vehicle operators and passengers. The user can interact with the services offered via data networks, video signals, or audio. The interaction can be performed with automated services or a live Customer Service Representative. Interactions with the customer service representatives can be through any combination of possible techniques, such as, live audio, live video, electronic messaging or email, instant messaging, and other techniques. These services can be offered by a number of entities including, vehicle manufacturers, vehicle dealers, vehicle service organizations, automobile or travel clubs, wireless carriers, travel service organizations, or other organizations. The services offered can be personalized to the occupants of the vehicle using a variety of information including, user profile information, history, location, paths traveled, time of day, day of week, or other information. In addition, the system can offer customized services based on information about the vehicle, including paths traveled, distance, service history, type of equipment on vehicle. These services can be accessed while a person is an occupant in a vehicle, while the person is not an occupant but are using a wireless or wired network equipped

Google Exhibit 1019
Page 00513

mobile device or while using a wired or wireless network desktop system, among other situations. Examples of these services include:

a. Location based marketing programs wherein occupants of the vehicle receive promotional offers from merchants along a route of travel. Occupants can query the system for offers and promotions for goods, services, or other information, along the travel route. The system may apply other available information to form a response, including, the user profile, history, location, or other information. The system can provide optimized interactive routing assistance to the vehicle operator. Alternatively, the system can provide interactive offers and promotions for goods and services along the route, or in advance of a particular trip. Promotions can be offered for goods and services can include but are not limited to, travel services, groceries, prepared foods, vehicle service, fuel, entertainment, or other goods and services.

b. Remote ordering and payment for goods and services. The system can interactively present the menu or product catalog using the list and table presentation capabilities of the invention. The system may facilitate remote ordering by using location information, customer preferences, customer order history, etc. The system can manage a secure payment wallet for the users. Voiceprints, spoken passwords, and non-speech security methods (i.e. PIN pad, etc.) can be combined to create the appropriate level of security.

c. Travel services for occupants of the vehicle. These services can include, directories of travel and entertainment services, or reservations for entertainment restaurants, hotels and other accommodation. The system may present directories, lists and menus using its interactive list and table presentation capabilities. The travel service capability can be used in conjunction with the remote ordering and payment capabilities and the dynamic interactive routing capability.

d. Answer specialized travel related questions in areas such as vehicle registration, taxes, safety laws, required inspections, weight

Google Exhibit 1019
Page 00514

limits, insurance coverage requirements, insurance policy provisions, or other areas.

6. The invention can provide an operator or other occupant of a vehicle with an interactive location sensitive shopping list or a location and time sensitive task reminder list using the natural language interface. Users can create the list while in the vehicle, while on foot using a mobile device, or at a fixed location using a handheld or desktop device. A user may grant permission to other users to add tasks or shopping items to their lists. Once in the vehicle, the system may provide occupants with routing assistance to optimize travel time and may provide reminders of items to be purchased and tasks to be completed as the vehicle comes in close proximity to a particular location, type of merchant or other service provider, or when a set time has been reached, among other criteria.

7. Automatic interactive dispatch and reporting for fleet vehicles having vehicle operators or other vehicle occupants that use the speech interface and/or non-speech interface to interact with these services. These services can include, dynamic optimal routing, inventory of parts and other materials, ordering of required parts and materials, work orders, receipt generation, payments, or other services.

8. Sales force automation, sales reporting, contact database management, calendar management, and call routing. The system may employ its interactive list and table presentation capabilities to supply catalog and pricing information or other information. These services can use local or network data. Additional services can include, memos, reminders, activity lists, or other information.

[046] Vehicle operators and other occupants can use the interactive natural language interactive interface of the invention to perform many types of information query, retrieval and presentation operations. Using the natural language interactive interface, users can modify the parameters of queries or specify the presentation formats for results. Data used to create a response can be from any combination of local and remote data sources. User specific data can be synchronized between systems fixed to one or more vehicles, mobile structures and desktop systems. Some examples of

Google Exhibit 1019
Page 00515

information query, retrieval and presentation applications for the invention include, but are not limited to the following:

1. White pages and yellow pages lookups to find, email addresses, telephone numbers, street addresses and other information for businesses and individuals. These services can be used in conjunction with other services, including remote ordering and payment, offers and promotions, mapping, and driving directions;

2. Management and access to personal address book, calendars and reminders for users;

3. Automatic telephone dialing, reading and sending emails, pages, instant messaging by voice, text or video and other communications control functions;

4. Selection, schedules and play list management for television, satellite broadcast, radio or other entertainment schedule. The available information can include reviews and other information on programming. The system may provide device control for users;

5. Weather information for the local area or other locations;

6. Stock and other investment information including, prices, company reports, profiles, company information, business news stories, company reports, analysis, price alerts, news alerts, portfolio reports, portfolio plans, or other information;

7. Local, national and international news information including headlines of interest by subject or location, story summaries, full stories, audio and video retrieval and play for stories;

8. Sports scores, news stories, schedules, alerts, statistics, background and history information, or other information;

9. The ability to subscribe interactively to multimedia information channels, including sports, news, business, different types of music and entertainment, applying user specific preferences for extracting and presenting information;

10. Rights management for information or content that is used or published;

Google Exhibit 1019
Page 00516

11. Horoscopes, daily jokes and comics, crossword puzzle retrieval and display and related entertainment or diversions;

12. Interactive educational programs using local and network material, with lesson material level that is set based on user's profile, location of the vehicle, planned route of the vehicle, planned activities during the trip and including, interactive multimedia lessons, religious instruction, calculator, dictionary and spelling, geographic information, instruction for specialized tasks planned during the trip, language training, foreign language translation, presentation of technical manuals, and encyclopedias and other reference material.

[047] It will be appreciated that the foregoing statements of the features of the invention are not intended as exhaustive or limiting, the proper scope thereof being appreciated by reference to this entire disclosure and reasonably apparent variations and extensions thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

[048] The invention will be described by reference to the preferred and alternative embodiments thereof, in conjunction with the drawings in which:

[049] Figure 1 is an overall block diagram of the system according to a first embodiment of the invention;

[050] Figure 2 is an overall block diagram of the system according to a second embodiment of the invention;

[051] Figure 3 is an overall block diagram of a mobile device according to an embodiment of the invention;

[052] Figure 4 is an overall block diagram of a fixed computer according to an embodiment of the invention;

[053] Figure 5 is an overall diagrammatic view of the interactive natural language processing system according to an embodiment of the invention;

[054] Figure 6 is a schematic block diagram showing the agent architecture according to an embodiment of the invention;

[055] Figure 7 illustrates a diagram for centrally maintaining context, according to one embodiment of the invention; and

Google Exhibit 1019
Page 00517

[056] Figure 8 illustrates a system diagram for an enhanced conversational voice user interface according to one embodiment of the invention.

DETAILED DESCRIPTION

[057] The following detailed description refers to the accompanying drawings, and describes exemplary embodiments of the invention. Other embodiments are contemplated and modifications may be made to the exemplary embodiments without departing from the spirit, functionality and scope of the invention. Therefore, the following detailed descriptions are not meant to limit the invention.

[058] According to one embodiment of the invention, a telematic natural language speech interface and non-speech interface are provided for use in mobile environments and telematic applications. The system, or portions thereof, may be used in vehicles, while on foot or at a fixed location such as an office or home, or at other locations. An overall block diagram of one embodiment of the invention is illustrated in Figure 1.

[059] A speech unit 128 and/or a keypad 14 may be integrally coupled to a mobile structure 10 or may be part of mobile devices 36, fixed home or office computer systems 44, or other devices. Mobile devices 36 may include mobile telephones, personal digital assistants, digital radios, compact disk players, navigation systems, or other mobile devices. The mobile devices 36 may be configured to integrate with set-top boxes, alarm clocks, radios, or other electronic components. The speech unit 128 and/or keypad 14 may be interfaced to a Telematics Control Unit (TCU) 28 through one or more data interfaces 26. According to some embodiments, the main speech-processing unit 98 may be embedded in one or more TCU 28. In some embodiments, the components of the speech unit 128 may also be distributed between one or more TCUs.

[060] A speech-processing unit may be built into mobile devices 36 and may be coupled with the data interfaces 26 though a wireless, or wired, handheld interface 20. Other user interface peripherals may be connected to the TCU through the data interfaces and may include displays 18, such as touch screen displays for displaying text, graphics and video; keypads 14 for receiving textual data input; video cameras 16 for receiving multimedia communications or conferences; a pointing device or stylus, or other user interface peripherals. Other devices may be connected to the TCU through the data interfaces including wide-area RF transceivers 24, navigation

Google Exhibit 1019
Page 00518

system components 22, or other devices. The navigation system may include several components such as, Global Positioning System (GPS) receivers or other radiolocation system receivers, gyroscopes or other inertial measurement equipment, distance measurement sensors, such as odometers, or other components. Radiolocation equipment may receive coded signals from one or more satellites or terrestrial sources 40. The one or more location service servers 48 may assist the navigation system. Other systems that can connect to the TCU through the data interfaces may include automotive control computers, digital control interfaces for devices such as media players or other electronic systems, measurement sensors, or other specialized electronic equipment.

[061] The control and device interfaces 30 may connect the TCU 28 to various devices 32 within the mobile structure 10. The control and device interfaces 30 may be used to execute local or remote commands from users of the natural language interface. In some cases, the control and device interfaces 30 may include specialized hardware for interacting with different types of devices. The hardware interfaces may include analog or digital signal interfaces for device control, along with analog or digital interfaces for measurements that may control devices 32. These interfaces may also include specialized software that encapsulates or abstracts specific behavior of the devices 32. The interface software may include one or more drivers that are specific to the hardware interface and to one or more agents. The domain agents may include the specialized software behavior and data required to control a particular device or class of devices. New or updated behavior may be added to the system by updating data managers that are associated with specific devices or class of devices. The devices 32 may include manual controls or manual overrides 34. For safety reasons, the control and device interfaces 30 may incorporate fail-safe systems that, for example, may verify operating limits before changing settings, ensuring that commands do not conflict with settings from manual controls, and will not, in some combination with other commands or device settings, create an unsafe situation. The software behavior and data that ensures safe operations may be included within the domain agent that is specific to the device or class of devices. Examples of devices and system that can be controlled through the control and device interfaces 30 include, power management systems, measurement sensors, door locks, window controls, interior temperature controls, shifting of the transmission, turn signals,

Google Exhibit 1019
Page 00519

lights, safety equipment, engine ignition, cruise control, fuel tank switches, seat adjustments, specialize equipment such as winches, lifting systems or loading system, and other systems.

[062] The wide-area RF transceiver 24 may communicate with one or more wide-area wireless networks 38, which may be connected to data networks 42, including the Internet, the Public Switched Telephone Network (PSTN) 42 or other data networks. The wide-area wireless networks can be of any suitable terrestrial or satellite based type. Mobile devices 36 may communicate with one or more local or wide-area wireless networks. Home or office systems 44, equipped with wired or wireless network interfaces, may communicate through the data networks or PSTN.

[063] According to one embodiment of the invention, data and agents may be stored and synchronized in mobile structures 10, mobile devices 36 and/or fixed systems 44 having one or more main speech-processing units 98. The synchronization between these different systems can occur on the wide area wireless network 38, the data network 42, through the handheld interface 20, or other local data connections. The synchronization may be performed automatically when any two or more of the computers are connected to these networks. Alternatively, the synchronization may be performed on demand under user control. The synchronization process attempts to determine which version of a data element or an agent are the newest or most up-to-date and propagates that element. Thus, synchronization is an incremental change processes. In some cases, a complete replacement of a database, a portion of the database or of one or more agents may be performed rather than performing a series of incremental updates.

[064] The wide-area wireless networks 38, the data networks 42 or PSTN, may connect mobile structures 10, mobile devices 36 and fixed computers 44 to one or more servers that provide one of more services. According to one embodiment of the invention, an interactive natural language user interface may be provided that supports a transfer of data or transmission of speech, text, video and other formats. For data centric applications, a standardized data transfer format may be used including, for example, Hypertext Markup Language over Hypertext Transfer Protocol (HTTP), Extensible Markup Language (XML), and other data formats or schemas, over HTTP or other transfer protocol, Electronic Data Interchange formats over a variety of transport protocols, etc. It will be understood that the exact configuration of the

Google Exhibit 1019
Page 00520

servers may be determined on a case-by-case basis with consideration being given to the exact combinations of services being offered, the service providers providing the services, the contractual relationships between the service provider, and other factors. The invention supports any suitable configuration. In each case these servers may themselves be distributed over one or more public or private networks. Some examples of servers that may be used to deliver these services, are given below:

1. One or more payment service providers 56 that supply payment capabilities to users. These payment services may include electronic wallet capabilities for one or more payment accounts and may include payment security information, payment account information, transaction histories, account balance information, and other information. Exemplary payment types include, stored value accounts, promotional accounts, credit accounts, telecommunications billing accounts, debit accounts, or other payment types that are accessed using online or offline methods. Payments can be computed in any manner including payment for a specific good or service, subscription payment, metered payment, or other payments. The payment services can be distributed in a number of ways. Examples of computers and servers used to store and process payment transaction information include, smart cards, main speech processing units 128, mobile devices 36, TCUs 28, fixed personal computers 44, payment gateways, payment servers 56, or other systems.

2. One or more Customer Relationship Management (CRM) systems 52 may supply any number of consumer and business customer services as discussed above. The CRM system can supply automated services or services that are partly or completely manual. For manual services one or more customer service representatives may use one or more service representative workstations 54. The CRM system and the service representative workstation may be connected to one or more data networks 42 or the PSTN. Any other servers may also have connections to one or more service representative workstations, which may be in common or independent of each other.

3. One or more specialized service servers 50 may support specialized consumer and business services.

4. One or more location service servers 48 may supply location information and location based services. The location data is used as a data

Google Exhibit 1019
Page 00521

input to the location services that may then be distributed in any suitable manner including, on main speech processing units 128, mobile devices 36, TCUs 28, fixed personal computers 44, other servers (i.e. 46, 50, 52, 56) or the one or more location services servers 48.

5. One or more emergency services servers 46 may supply both public and private emergency services to the users.

[065] Main speech processing unit 98, speech unit 128 and keyboard 14 may be distributed in mobile devices 36 in a number of ways. For example, these units may be attached to the mobile devices 36 as independent components or as a single integrated component. In another embodiment of the invention, some or all of the main speech processing unit 98 and speech unit 128 may be embedded in one or more of the TCUs 28, mobile devices 36, fixed computer systems 44, or other devices.

[066] Figure 2 illustrates an alternative block diagram of a second embodiment of the invention. In this embodiment, the main speech processing unit 98 and speech unit 128 are external to the TCU 28. These components may be housed in one or more packages or may be included in a single integrated package.

[067] In all other respects, the second embodiment resembles the first embodiment illustrated in Figure 1. It will be understood that the exact distribution and packaging of the main speech processing unit 98 and speech unit 128 may be determined by the details of the deployment situation and should not change the functionality, capabilities or spirit of the invention in any way.

[068] Figure 3 illustrates a block diagram of one embodiment of the mobile devices 36 that include main speech processing unit 98 and speech unit 128 embedded in the mobile device 36, mobile telephone or other mobile device. The main speech processing unit may interface to one or more processing units 70 located in the mobile device 36. The processing units 70 may include one or more central processing units, one or more data and address busses, data interfaces, volatile memory, or other components. The processing units 70 may use one of more types of nonvolatile memory 80 for software and data storage. Suitable types of nonvolatile memory 80 may include flash memory and hard disk drives. In some embodiments, the main speech processing unit 98 may be integrated with the one or more processing units 70.

[069] According to one embodiment, users may interact with the mobile device 36 through the speech unit 128, the keypad 74 or keyboard, a display 72 that displays

Google Exhibit 1019
Page 00522

text, graphics, video, or other peripheral. In some embodiments the display may be a touch screen type. Alternatively, a pointing device (not shown) may be used, among other devices.

[070] The mobile device 36 may connect to one or more wired or wireless wide-area or local-area networks through one or more interfaces. A wide-area network transceiver 78 may connect to the wide-area wireless network 38 or the data network 42 using a wireless or wired connection, including an IP connection, a dial-up PSTN network connection or other connections. The local-area network transceiver 76 may connect to wired or wireless local area networks. These networks may include the handheld interface 20 or connections to fixed computer systems 44. In both mobile device environments and fixed computer environments, communications between humans and machines may not provide accurate results at least because natural language commands may be incomplete in their definition. The occurrence of inaccurate results may be reduced by leveraging context of previous utterances, knowledge of the domain and/or the user's history of interests and preferences or other factors.

[071] According to one embodiment of the invention, processing may be performed at the mobile devices 36. Alternatively, the processing may be performed on the server side. In yet another embodiment, the processing may occur on both the mobile devices 36 and the server side. The commands may be processed on-board the mobile devices 36 to enable the mobile devices 36 to control themselves and/or to control other mobile devices 36, fixed computers 44, mobile telephones, or other devices. Additionally, mobile devices 36 may track context. According to another embodiment of the invention illustrated in FIG. 7, a context manager 702 may be centrally maintained to allow input to and output from multiple mobile devices 36. Each of the mobile devices may communicate with the context manager 702 to register via registration module 712, wherein registration may indicate events that the mobile devices 36 may subscribe to. The context manager 702 may receive input in Context XML form, for example. The other registered mobile devices 36 may be informed of context changes through context tracking module 714, thereby synchronizing the context across the registered mobile devices 36. According to one embodiment of the invention, registered mobile devices 36 may be added or removed.

Google Exhibit 1019
Page 00523

The registered mobile devices 36 may be dynamic link libraries (DLLs), object code, or other data that may be specific to the mobile device 36.

[072] According to yet another alternative embodiment of the invention, context information may be determined from a command or request that is presented as a text message and/or a command or request that is presented as a verbal utterance and processed using a multi-pass automatic speech recognition module that transcribes the verbal utterance to a text message. The multi-pass automatic speech recognition module may use a dictation grammar or a large vocabulary grammar, among other resources, to transcribe the verbal utterance into a text message. On platforms where a dictation grammar is not available, the multi-pass automatic speech recognition may use a virtual dictation grammar that uses decoy words for out-of-vocabulary words. Decoy words may include utility words, nonsense words, isolated syllables, isolated distinct sounds and other decoy words.

[073] According to one embodiment of the invention, the text message may be searched for a particular character, group of characters, word, group of words, and other text combinations. The text combination may be compared against entries in a context description grammar that is associated with each agent 106. If a match is identified between an active grammar in the context description grammar and the command and/or request, then the match may be scored. The agents 106 may be ranked based on the determined score. In generating an aggregate response from the one or more responses received from the agents 106, the ordering of the responses from the individual agents may be determined based on the rank of agents 106. The aggregate response may be generated by a response generator module. Agents 106 may update a context stack, that includes an ordered list of command contexts, to enable follow-up requests.

[074] According to another embodiment of the invention, if a match is not found, or only a partial match is found, between the text message and active grammars, then a knowledge-enhanced speech recognition system may be used to semantically broaden the search. The knowledge-enhanced speech recognition system may be used to determine the intent of the request and/or to correct false recognitions. The knowledge-enhanced speech recognition may access a set of expected contexts that are stored in a context stack to determine a most likely context. The knowledge-enhanced speech recognition may use context specific matchers that are able to

Google Exhibit 1019
Page 00524

identify context such as time, location, numbers, dates, categories (e.g., music, movies, television, addresses, etc.) and other context. The matching may be performed by comparing a character, group of characters, a word, group of words, and other text combinations. Alternatively, or in addition to text based matching, the matching may be performed using phonetic matching, among other techniques. The results of any match may be used to generate a command and/or request that is communicated to agents 106 for additional processing. According to one embodiment of the invention, non-speech interface 114 may show system, state and history information in a more concise manner than is possible through the speech interface. Non-speech interface 114 may be accessed to create or extend capabilities of agents 106. These operations may include scripting of agents, adding data to the agent or databases 102 used by the agent, adding links to information sources, among other operations.

[075] According to another embodiment of the invention, mobile devices 36 may be configured to allow speech annotations of objects stored thereon. The objects may include photographs, calendar entries, email messages, instant messages, phonebook entries, voice mail entries, digital movies or other objects. Mobile devices 36 may transcribe the speech annotations to textual annotations and store the textual annotations with the object. Alternatively, mobile devices 36 may be configured to enable users to input non-speech annotations, such as textual descriptions, that are stored along with the objects.

[076] According to one embodiment of the invention, the annotated objects may be stored on a server side, a client side, a combination of server side and client side, or according to other configurations. The invention further contemplates collaboratively exchanging and sharing the annotated objects among distributed workgroups that may include centralized servers having shared workspaces for providing common object storage and retrieval facilities. The shared workspaces may be implemented on the centralized servers and may be accessed from different platforms using the mobile devices 36. The system may include a peer-to-peer system for accessing the annotated objects.

[077] According to one embodiment of the invention, the non-speech annotations may be classified and searched. In an alternative embodiment, the speech annotations may be classified and searched. However, classifying and searching speech

Google Exhibit 1019
Page 00525

annotations may be considerably more difficult than classifying and searching non-speech annotations.

[078] According to another embodiment of the invention, the textual annotations and/or textual descriptions that are associated with the objects may be stored as metadata, thereby enabling searching for the objects using the metadata. The metadata may include GPS information, environmental information, geographic information, or other information. For example, proximity to famous landmarks may be determined using GPS information, environmental information, geographic information, or other information, and this information may be integrated into the metadata associated with the objects. According to one embodiment of the invention, GPS coordinates may be stored in the metadata associated with the objects and users may search for selected objects based on the GPS coordinates. A user may provide a speech command such as "show me all the photos of Greece." In this case, the system would limit the type of object to photographs and would determine the GPS coordinates of Greece. The system would then search the metadata for objects that correspond to photographs and that also satisfy GPS coordinates for Greece. According to another embodiment of the invention, including GPS coordinates in the metadata of objects enables post-processing of the objects based on the GPS coordinates. For example, the objects may be subjected to an initial sort based on generalized GPS coordinates that are stored in the metadata and may be subjected to additional sorting based on more particular criteria for the GPS coordinates. Thus, a user may initially search for object metadata that corresponds to a location near a famous landmark and may use image matching to label the objects with the searchable metadata (i.e., textual descriptions). For example, using the GPS coordinates stored in the metadata, users may first determine that photographs were taken at the Jefferson Memorial and may use this information to label the photographs with searchable metadata (i.e., textual descriptions) including "photo of the Jefferson Memorial." One of ordinary skill in the art will readily appreciate that metadata may include various types of information and may be searched using the various types of information.

[079] According to one embodiment of the invention, the textual annotations may be communicated using a short message service on the mobile telephone or other device. Short message service is a text message service that enables sending and receiving of

Google Exhibit 1019
Page 00526

short textual messages. The textual messages may be stored at data centers for forwarding to intended recipients. Other configurations may be used.

[080] According to another embodiment of the invention, the mobile devices 36 may support multi-modal communications that enable displaying of non-speech search results on a graphical interface and receipt of speech commands to provide a follow-up search, among other configurations. For example, the user may be presented with textual search results corresponding to a name of a famous person and the user may provide a speech command to find a biography of the famous person. The system may maintain the context of the textual search results to find the biography associated with the famous person. By contrast, known systems may perform a follow-up search on the term "biography" and may present a dictionary definition of the term "biography."

[081] Figure 4 illustrates an embodiment that uses a fixed computer 44. In some embodiments, the main speech processing unit 98 and speech unit 128 may be embedded into the fixed computer 44. The main speech processing unit may interface with one or more processing units 84 that are associated with the fixed computer. The processing units may include one or more central processing units, one or more data and address busses, data interfaces, volatile memory, or other components. The processing unit may use one of more types of nonvolatile memory 94 for software and data storage. Suitable types of nonvolatile memory include, for example, flash memory and hard disk drives. In some embodiments, the main speech processing unit 98 may be integrated with the one or more processing units 84. According to one embodiment of the invention, processing may be performed at the fixed computers 44. The commands may be processed on-board to enable the fixed computers 44 to control themselves and/or to control other fixed computers 44, mobile devices 36, mobile telephones, or other devices. Additionally, fixed computers 44 may track context as described above.

[082] According to another embodiment of the invention, fixed computers 44 may be configured to allow verbal annotations of objects stored thereon. Fixed computers 44 may transcribe the verbal annotation to text and store the textual annotation with the object. Alternatively, fixed computers 44 may be configured to enable users to manually input textual descriptions that are stored along with the objects. According to one embodiment of the invention, the textual annotations and/or textual

Google Exhibit 1019
Page 00527

descriptions may be classified and searched. In an alternative embodiment, fixed computers 44 may classify and search the verbal annotations rather than textual annotations. However, classifying and searching verbal annotations may be considerably more difficult than classifying and searching textual annotations and/or textual descriptions.

[083] In another embodiment of the invention, users may interact with fixed computers 44 using speech unit 128, the keyboard 88 or keypad, a display 86 for displaying text, graphics, video, or other peripherals. According to some embodiments of the invention, the display may be a touch screen type. Alternatively, a pointing device (not shown) may be used, along with other devices. Fixed computers 44 may be coupled to one or more wired or wireless wide-area or local-area networks through one or more interfaces. A wide-area transceiver 92 may connect to the wide-area wireless network 38 or the data network 42, using a wireless or wired connection, including an IP network, a dial-up PSTN network connection, or other connections. The local-area network transceiver 90 may connect to wired or wireless local area networks. These networks may include connections to mobile devices 36.

[084] In order for devices to properly respond to requests and/or commands that are submitted in a natural language form, machine processable requests and/or algorithms may be formulated after the natural form questions or commands have been parsed and interpreted. Algorithms describe how the machines should gather data to respond to the questions or commands. Depending on the nature of the requests or commands, there may not be a simple set of requests and/or algorithms that will return an adequate response. Several requests and algorithms may need to be initiated and even these requests and algorithms may need to be chained or concatenated to achieve a complete response. Further, no single available source may contain the entire set of results needed to generate a complete response. Thus, multiple requests and/or algorithms, perhaps with several parts, may be generated to access multiple data sources that are located both locally or remotely. Not all of the data sources, requests and/or algorithms may return useful results or any results at all. Useful results that are returned are often embedded in other information and may need to be extracted from the other information. For example, a few key words or numbers may need to be "scraped" from a larger amount of other information in a text string, table, list, page,

Google Exhibit 1019
Page 00528

video stream or other information. At the same time, extraneous information including graphics or pictures may be removed to process the response. In any case, the multiple results must be evaluated and combined to form the best possible response, even in cases where some requests do not return useful results or fail to produce results entirely. In cases where the command is determined to be ambiguous or the result is inherently subjective, determining the results to present in the response is a complex process. Finally, to maintain a natural interaction, responses should be returned to the user rapidly. Managing and evaluating complex and uncertain requests, while maintaining real-time performance, is a significant challenge.

[085] The invention provides a complete speech-based command generation, information query, retrieval, processing and presentation environment or a combination of speech-based and non-speech-based command generation, information query, retrieval, processing and presentation environment for telematic applications. In addition, the invention may be useful for controlling the system itself and/or external devices. This integrated environment makes maximum use of context, prior information and domain and user specific profile data to achieve a natural environment for one or more users submitting requests or commands in multiple domains. Through this integrated approach, a complete speech-based natural language command, algorithm and response environment or a combination of speech-based and non-speech-based command, algorithm and response environment may be created.

[086] The telematic natural language interface may be deployed as part of, or a peripheral to a TCU or other mobile devices 36, as part of a mobile device interfaced to vehicle computers and other mobile system through wired, wireless, optical, or other types of connections or fixed computers interfaced to the vehicle computers or other systems through a combination of wired, wireless, optical and/or other types of connections. Alternatively, the components of the interactive natural language telematic interface may be distributed in any suitable manner between these multiple computing platforms.

[087] Figure 5 illustrates one exemplary schematic diagram view of a system 90 according to an embodiment of the invention. The system 90 may include a main unit 98, speech units 128, and multi-modal devices 155. In an alternative embodiment, the system 98 may include completely separate systems for main unit 98, speech units

Google Exhibit 1019
Page 00529

128 and multi-modal devices 155. The event manager 100 may mediate interactions between components of main unit 98. The event manager 100 provides a multi-threaded environment allowing the system 98 to operate on multiple commands or questions from multiple user sessions, without conflict, and in an efficient manner to maintain real-time response capabilities.

[088] FIG. 8 illustrates one exemplary embodiment of an enhanced conversational user interface 800 that receives multi-modal input from users 802 and communicates with a conversational speech analyzer 804. The conversational speech analyzer 804 may be coupled to several components, including a general cognitive model 806, an environmental model 808, a personalized cognitive model 810, and an adaptive misrecognition analysis engine 812 or other components. According to one embodiment of the invention, the conversational speech analyzer 804 may include one or more speech recognition engines that transcribe utterances to textual messages. The transcription may be performed using one or more conversational grammars, a context description grammar 112, or other transcription methods. According to one embodiment of the invention, data that is entered in a text format may be merged with data that is transcribed to a textual format from the utterance. Conversational speech analyzer 804 also may include a semantic knowledge-based model that analyzes the textual message and detects command components.

[089] According to one embodiment of the invention, personalized cognitive model 810 is a model derived from a user's interaction pattern with the system and may be used to predict what actions the user may take next in time, thus assisting with speech recognition and/or question or command recognition. Personalized cognitive model 810 may track actions performed by the user. When the system is attempting to predict user behavior, the personalized cognitive model may be consulted first. The system may have multiple personalized cognitive models, wherein one may correspond to each user.

[090] According to another embodiment of the invention, general cognitive model 806 is a statistical abstract that corresponds to interaction patterns with the system for multiple users. Data stored within general cognitive model 806 may be used to predict a user's next action, thus assisting with speech recognition and/or question or command recognition. The general cognitive model 806 may also track what actions

Google Exhibit 1019
Page 00530

a particular user has performed and may be used when the user interacts with the system in a way that is not handled in the personalized cognitive model.

[091] According to one embodiment of the invention, the environmental model 808 may include information associated with the user's environment and surroundings. The information may include the type of environment that a user is in (e.g., quiet or noisy); details of a microphone and/or speaker system; the user's current global position and movement, as may be determined by GPS; current system status, such as what song/movie is playing, is the system in the midst of retrieving something, or other system status; details on all voice-enabled devices in the immediate vicinity, such as a presence of a voice-enabled TV, stereo, and DVD player in the same room; a user's credit card information, such as numbers and current balances, wherein the user may ask a mobile telephone to download and pay for a video and the system may respond that there are insufficient funds; or other information. The information may be accessed to invoke a context, domain knowledge, preferences, and/or other cognitive qualities that enhance the interpretation of questions and/or commands.

[092] The conversational speech analyzer 804 may also access the general cognitive model 806 and/or the personalized cognitive model 810 to further refine a context, domain knowledge, preferences, and/or other cognitive qualities to enhance the interpretation of questions and/or commands. Based on information received from general cognitive model 806, environmental model 808 and/or the personalized cognitive model 810, the system may enhance responses to commands and questions by including a prediction of user behavior.

[093] Adaptive Misrecognition Analysis Engine 812 may analyze and store textual messages, including transcribed utterances, that are identified as being unrecognized or incorrectly recognized by conversational speech analyzer 804. Upon a determination that text is unrecognized, the system may generate an unrecognized event. For example, an unrecognized event may result from not finding a match to text and/or the transcribed utterance.

[094] According to one embodiment of the invention, the system may implement one or more techniques to determine that textual messages are incorrectly recognized. For example, the user may command the system to play a specific song title and the system may misrecognize the requested song title and provide a song having a different title, the system may misrecognize an incorrect song title and

Google Exhibit 1019
Page 00531

provide a song having a different than requested title, the system may misrecognize a correct song title and provide a song having an invalid song title, among other variations. When the system misrecognizes a request, the user typically provides immediate feedback such as, overriding the command in a time shorter than the expected execution time of the command, repeating the original request, issuing a stop command, or taking other action, wherein the action may be presented verbally, non-verbally by pushing a selected button on a cell phone or remote control, or both, among other configurations. According to one embodiment of the invention, the system may detect the user action and may prompt the user to re-phrase the request to enable the system to modify words in the query. The user's actions may be analyzed in a statistical model to determine a frequency occurrence of misrecognitions for particular commands, with the results being used to update the corresponding personalized cognitive model 810.

[095] According to another embodiment of the invention, the conversational speech analyzer 804 may access the personalized cognitive model 810 to proactively select a next best (or nth best) match for the received text. A match may be confirmed by user action that include the user not immediately canceling the command or taking other action. The misrecognitions may also be analyzed to potentially determine personalized tuning parameters for the speech recognition components of the system. For example, the system may, over time, tune the speech recognition engine to increase recognition by analyzing how the speech recognition engine misrecognizes the utterance.

[096] The event manager 100 may mediate interactions between other components of the invention. The event manager can provide a multi-threaded environment allowing the system to operate on multiple commands or questions from multiple user sessions without conflict and in an efficient manner, maintaining real-time response capabilities.

[097] Agents 106 may include a collection of grammars, criteria handlers, and algorithms that are accessible to respond to a set of requests and/or commands. Agents 106 may further contain packages of both generic and domain specific behavior for the system 98. Agents 106 may use nonvolatile storage for data, parameters, history information, and locally stored content provided in the system databases 102 or other local sources. One or more user profiles 110 may be provided

Google Exhibit 1019
Page 00532

that include user specific data, parameters, and session and history information for determining the behavior of agents 106. One or more personality modules 108 may be provided in a data determining system that include personality characteristics for agents. The update manager 104 manages the automatic and manual loading and updating of agents 106 and their associated data from the Internet 146 or other network through the network interface 116.

[098] According to one embodiment of the invention, the speech-based interface for the system 90 may include one or more speech units 128. Speech units 128 may include one or more microphones, for example array microphone 134, to receive the utterances from the user. The speech received at the microphone 134 may be processed by filter 132 and passed to the speech coder 138 for encoding and compression. In one embodiment, a transceiver module 130 may transmit the coded speech to the main unit 98. Coded speech received from the main unit 98 is detected by the transceiver 130, then decoded and decompressed by the speech coder 138 and annunciated by the speaker 136.

[099] According to one embodiment of the invention, the non-speech-based interface for the system 90 may include one or more multi-modal devices 155 that may include mobile devices, stand alone or networked computers, personal digital assistances (PDAs), portable computer devices, or other multi-modal devices.

[0100] The speech units 128, multi-modal devices 155 and the main unit 98 may communicate over a communication link. The communication link may include a wired or wireless link. According to one embodiment, the communication link may comprise an RF link. The transceiver 130 on the speech unit may communicate coded speech data bi-directionally over the communication link with the transceiver 126 on the main unit 98. According to another embodiment, the RF link may use any standard local area wireless data protocols including the IEEE 802.11, Bluetooth or other standards. Alternatively, an infrared data link conforming to any suitable standard such as IrDA or other infrared standards can be used. In an alternative embodiment, wires may connect the speech unit 128 and the main unit 98, eliminating the need for one speech coder 138. Other wired or wireless analog or digital transmission techniques can be used.

[0101] According to one embodiment of the invention, coded speech received at the transceiver 126 on the main unit 98 is passed to the speech coder 122 for decoding

Google Exhibit 1019
Page 00533

and decompression. The decoded speech may be processed by the speech recognition engine 120 using the context description grammar module 112, among other information. Any recognized information may be processed by the parser 118, which transforms information into complete algorithms and questions using data supplied by knowledge agents. Knowledge agents may be selected from the plurality of agents 106 using a grammar stack, wherein the knowledge agents provide information for generating a response to the question or command. The knowledge agents may then process the commands or questions by creating requests that are submitted to local databases 102 or submitted through the network interface 116 to external data sources over the Internet 146 or external other networks. Algorithms typically result in actions taken by the system 90 itself (i.e., pause or stop), or to a remote device or data source (i.e., download data or program, or control a remote device), through the network interface to the Internet or other data interface.

[0102] According to one embodiment of the invention, knowledge agents may return results of questions as responses to users. The responses may be created using the results of information requests, the system personality 108, the user preferences, other data in the user profile 110, and/or other information. Agents 106 may present these results using the speech unit 128. The agents 106 may create a response string, which is sent to the text to speech engine 124. The text to speech engine 124 may generate the utterances, which may be encoded and compressed by the speech coder 122. Once coded, the utterances are transmitted from the main unit 98 by the transceiver 126 to the transceiver 130 on the speech unit 128. The utterance is then decoded and decompressed by the speech coder 138 and output by the speaker 136. Alternatively, agents 106 may present the results using multi-modal devices 155.

[0103] The non-speech interface 114 may be part of, or separate from, the multi-modal devices 155 and may be used as a substitute or complement to the speech interface. For example, non-speech interface 114 may be used to present and interact with non-speech (e.g., graphical or tabular) information in a manner more easily understood by the user. According to one embodiment of the invention, multimodal support may be provided to maintain the context during both voice interaction and interaction through the non-speech interface 114. In one exemplary embodiment, a communication channel may be opened between multimodal devices 155 and the main user interface system 90 to allow multimodal devices 155 to input text

Google Exhibit 1019
Page 00534

commands and questions. According to another embodiment of the invention, multimodal devices 155 may send a string of text or keywords for a command or question. The main interface system 90 may synchronize the context between multimodal device 155 and the speech units 128. In order to send a response to the corresponding device, the main user interface system 90 may track where the input came from so that the response may be sent to a TTS or multi-modal device 155.

[0104] According to one embodiment of the invention, non-speech interface 114 may show system, state and history information in a more concise manner than is possible through the speech interface. Non-speech interface 114 may be accessed to create or extend capabilities of agents 106. These operations may include scripting of agents, adding data to the agent or databases 102 used by the agent, adding links to information sources, among other operations.

[0105] According to another embodiment of the invention, system 90 may include different types of agents 106. In some embodiments of the invention, generic and domain specific behavior and information may be organized into domain agents 156. The system agent, on the other hand, may provide default functionality and basic services. The domain specific agents may provide complete, convenient and re-distributable packages or modules for each application area. In other words, the domain agents may include data that is needed to extend or modify the functionality of the system 90 in a current or new domain. Further, domain agents and their associated data can be updated remotely over a network as new behavior is added or new information becomes available. Domain agents may access a plurality of sources that may provide various services. Domain agents may use the services of other, typically more specialized, data managers and system agent. Agents are distributed and redistributed in a number of ways including on removable storage media, transfer over networks or attached to emails and other messages. The invention may provide license management capability allowing the sale of data managers by third parties to one or more users on a one time or subscription basis. In addition, users with particular expertise can create data managers, update existing data managers by adding new behaviors and information and making these data managers available to other users. A block diagram of the agent architecture according to an embodiment of the invention is shown in Figure 6.

Google Exhibit 1019
Page 00535

[0106] Agents 106 may receive and return events to the event manager 100. Both system agents 150 and domain agents 156 may receive questions and commands from the parser 118. Based on keywords in the questions and commands and the structures of the questions and commands, the parser may invoke selected agents. Agents use the nonvolatile storage for data, parameters, history information and local content provided in the system databases 102.

[0107] According to one embodiment of the invention, when the system starts-up or boots-up, the agent manager 154 may load and initialize the system agent 150 and the one or more domain agents 156. Agent manager 154 includes knowledge of agents 106 and maps agents 106 to the agent library 158. At shutdown, the agent manager may unload the agents 106. The agent manager 154 also performs license management functions for the domain agents 156 and content in the databases 102.

[0108] The system agent 150 manages the criteria handlers 152 that handle specific parameters or values (criteria) used to determine context for questions and commands. According to one embodiment of the invention, criteria handlers 152 include parsing routines that are specialized to recognize particular parts of speech, such as times, locations, movie titles, and other parts of speech. The criteria handlers 152 may identify matching phrases and extract semantic attributes from the phrases. Both the system agent 150 and the domain agents 156 may use the criteria handlers 152. The various domain agents 156 may use the services of the system agent 150 and of other, typically more specialized, domain agents 156. The system agent 150 and the domain agents 156 may use the services of the agent library 158, which contains utilities for commonly used functions. According to one embodiment of the invention, the agent library may be a dynamic link library that implements one or more agents. The agent library may include utilities for text and string handling, network communications, database lookup and management, fuzzy and probabilistic evaluation, text to speech formats, and other utilities.

[0109] Domain agents 156 may be data-driven, scripted or created with compiled code. A base of generic agent may be used as the starting point for data-driven or scripted agents. Agents created with compiled code are typically built into dynamically linkable or loadable libraries. Developers of agents can add new functionality to the agent library 158 as required. Details of agent distribution and update, and agent creation or modification are discussed in sections below.

Google Exhibit 1019
Page 00536

[0110] According to another embodiment of the invention, capabilities are provided to distribute and update system agents 150, domain agents 156, agent library 158 components, databases 102, and context description grammar 112 over wireless or wired networks 136, including IP networks and dial-up networks using the update manager 104. The network interface 116 may provide connections to one or more networks. The update manager 104 may also manage the downloading and installation of core system updates. The agent manager 154 may perform license management functions for the domain agents and the databases. The update manager 104 and agent manager 154 may perform these functions for all agents and database content including, agents and content available to all users or agents and/or content available to selected users. Examples of agent and database components added or updated on a periodic basis include:

- agents for new domains;
- additional domain knowledge for agents;
- new keywords for a domain, which can include names of politicians, athletes, entertainers, names of new movies or songs, etc. who have achieved recent prominence;
- links to a preferred set of information sources for the domains covered including links for, entertainment, news, sports, weather, etc.;
- updates to domain information based on, for example, changes to tax laws, company mergers, changing political boundaries;
- updates to content, including dictionaries, encyclopedias and almanacs; and
- other content and database components.

[0111] When a user requires or selects a new domain agent 156 or database element 102, the update manager 104 may connect to their source on the network 146 though the network interface 116, download and install the agent or data. To save system resources and to comply with any license conditions, the update manger 104 may uninstall agents that are no longer in use. In one embodiment of the invention, the update manager 104 may periodically query one or more sources of the licensed agents and database components to locate and download updates to agent executables, scripts or data as they become available. Alternatively, the agent sources may initiate

Google Exhibit 1019
Page 00537

the downloading of agent updates of the registered or licensed agents to the update manager as they become available.

[0112] The agent manager 154 may provide license management clients that are capable of executing most any license terms and conditions. When a particular agent 106 and/or database element 102 is selected based on a submitted command, the agent manager 154 verifies that the use of the agent or data element is within the allowed terms and conditions, and if so, invokes the agent or allows access to the data element. License management schemes that can be implemented through the agent manager 154 include outright purchase, subscription for updates, one time or limited time use. Use of shared agents and data elements (such as those down-loaded from web sites maintained by groups of domain experts) may also be managed by the agent manager 154.

[0113] If questions or commands do not match an agent that is currently loaded on the system, the agent manager 154 may search the network 146 through the network interface 116 to find a source for a suitable agent. This process may be triggered, for example, when a query is made in a domain for which an agent is not available, when a new device is added to a mobile structure, or when the behavior of a mobile device is updated. Once located, the agent can be loaded under the control of the update manager 104, within the terms and conditions of the license agreement, as enforced by the agent manager.

[0114] New commands, keywords, information, or information sources can be added to any domain agent 156 by changing agent data or scripting. These configuration capabilities may allow users and content developers to extend and modify the behavior of existing domain agents 156 or to create new domain agents 156 from a generic agent without the need to create new compiled code. Thus, the modification of the domain agents 156 may range from minor data-driven updates by even the most casual users, such as specifying the spelling of words, to development of complex behavior using the scripting language as would typically be done by a domain expert. The user can create and manage modifications to domain agents 156 through speech interface commands or using non-speech interface 114. User-specific modifications of domain agents 156 are stored in conjunction with the user's profile 110 and accessed by the domain agent 156 at run-time.

Google Exhibit 1019
Page 00538

[0115] The data used to configure data driven agents 156 may be structured in a manner to facilitate efficient evaluation and to help developers with organization. This data is used not only by the agent, but also in the speech recognition engine 120, the text to speech engine 124, and the parser 118. Examples of some major categories of data include:

1. Content packages may include questions or commands. Each command or question or group of commands or questions may include contexts used for creation of one or more requests. The domain agent 156 may pass a context description grammar expression to the parser 118 for evaluation of a context or question. An initial or default context is typically supplied for each command or question. The command or question includes a grammar for the management and evaluation of the context stack.
2. Parameters and other operating data on devices that are under control of the natural language speech interface. The domain agent 156 may use this data and parameters to determine how to execute a command, how to formulate the command string for the parser 118, determine if the command is feasible, and determine if the command can be executed within safety and operating limits.
3. Page lists or pointers to other local or network content sources. For each page or content source, there may be a pointer (e.g. URL, URI, or other pointer) to the page or source. Each page may have specific scraping information used to extract the data of interest. The scraping information may include, for example, matching patterns, HTML or other format parsing information.
4. A response list, determining the response of the domain agent 156 to a particular command or question given the context, the user profile and the information retrieved. Responses may include diagnostic error messages or requests for more information if the question or command cannot yet be resolved from the known information. Responses can be based on or be dependent on thresholds or probabilistic or fuzzy weights for the variables.
5. Substitution lists that include variable substitutions and transformations, often applied by the agents 150, 156 in the formatting of requests and results. For example, a stock domain specific agent 156 would

Google Exhibit 1019
Page 00539

use a substitution list of company trading symbols, company names and commonly used abbreviations. Substitutions and transformations can be performed on commands and questions to create precise queries, which can be applied against one or more information sources or to results for creating more meaningful output to the user. Substitution lists also include information for optimally dealing with structured information, such as HTML formatted page parsing and evaluation.

6. Personalities used for responses. Personalities can be constructed by combining multiple traits in a weighted manner. Weights can be specified for each agent's domain area to create one or more specific personalities. Examples of personality traits include, sarcasm, humor, irritation, and sympathy, and other traits.

7. Public and user specific parameters for sources, substitutions, transformations, variables or criteria. The public parameter lists are part of the agent package 156. The user specific parameters are included in the user profile 110.

[0116] Commands and questions are interpreted, requests formulated, responses created and results presented can be based on the user's personal or user profile 110 values. Personal profiles may include information specific to the individual, their interests, their special use of terminology, the history of their interactions with the system, domains of interest, or other factors. The personal profile data can be used by the agents 106, the speech recognition engine 120, the text to speech engine 124, and the parser 118. Preferences can include, special (modified) commands, past behavior or history, questions, information sources, formats, reports, alerts or other preferences. User profile data can be manually entered by the user and/or can be learned by the system based on user behavior. User profile values can include: 1) spelling preferences; 2) date of birth for user, family and friends; 3) income level; 4) gender; 5) occupation; 6)location information such as, home address, neighborhood, and business address, paths traveled, locations visited; 7) vehicle type or types; 8) vehicle operator certifications, permits or special certificates; 9) history of commands and queries; 10) telecommunications and other service providers and services; 11) financial and investment information; 12) synonyms (i.e., a nick name for someone,

Google Exhibit 1019
Page 00540

different terms for the same item); 13) special spelling; 14) keywords; 15) transformation or substitution variables; 16) domains of interest; or, 17) other values.

[0117] End users may use the data driven agent 156 extension and modification facilities and values stored in user profiles 110 to create special reports, packages of queries, alerts and output formats. A single alert or report can be configured to use multiple data sources and other variables (i.e., time, location, measured value) value to determine when alerts should be sent. For example, an alert can be generated by sampling a stock price every 15 minutes and sending an alert if the price drops below some value. In another example, an alert may be generated when a particular condition or combination of conditions occur for the vehicles. Alerts and reports can be directed to a local or remote output.

[0118] To create a report, the user may first specify a set of commands or questions. Next, the user can create or select a format for the report. Finally the user may name the report. A report can have variable parameters. For example, a user may create a company stock report and execute the report by stating its name and the company name, which gives the user selected information in a specified format for that company. In another example, a user can create a "morning" report, which presents selected multimedia information from different sources (news, sports, traffic, weather) in the order and formats desired. In yet another example, the user can create a report on the status of one or more vehicle systems. Alerts and reports can be created using only voice commands and responses, commands and responses through the graphical user interface 114, or a combination of the two. Reports can be run locally or remotely with respect to the vehicle. To create a report, alert, or other specialized behavior, the user performs a number of steps including: 1) specifying the command to run a report or alert; 2) specifying the question or questions, including keywords, used for a query; 3) setting the criteria for running the report such as on command or when a particular condition is met; 4) defining preferred information sources; 5)defining preferences for order of result evaluation by source, value, and other parameters; 6) specifying the presentation medium for a report or alert, such as an email, the text to speech engine, a message to a pager, or a text and graphics display; and, 7) specifying the preferred format for the report, such as information to be presented, order of information to be presented, preferred abbreviations or other variable substitutions.

Google Exhibit 1019
Page 00541

[0119] Filtering and noise elimination may be important in facilitating the various functionalities of system 90 and may improve operation in noisy mobile environments. Recognition and parsing of the user's speech is implemented with good signal to noise ratio at the input to the speech recognition engine 120. To provide acceptable results, a set of acoustic models, an array microphone 134, a filter 132, or other components, may be employed. If a good signal to noise ratio cannot be achieved, a noise identification algorithm may be used and the appropriate acoustic model, for example, one that has been trained in conditions similar to the identified noise, may be selected. According to one embodiment of the invention, the microphone array, filters and speech coder 138 are physically separated from the main unit 98 into a speech unit 128, and connected using a wireless link. Since bandwidth on a wireless connection is at a premium, the speech coder dynamically adapts the digitization rate and compression of the captured speech.

[0120] Some embodiments of the invention may use one or more arrays of microphones 134 to provide better directional signal capture and noise elimination than can be achieved with a single microphone. The microphone array can be one-dimensional (a linear array) or two-dimensional (a circle, square, triangle or other suitable shape). The beam pattern of the array can be fixed or made adaptive though use of analog or digital phase shifting circuitry. The pattern of the active array is steered to point in the direction of the one or more users speaking. At the same time, nulls can be added to the pattern to notch out point or limited area noise sources. The use of the array microphone also helps reduce the cross talk between output from the text to speech engine 124 through the speaker 136 or from another user talking and detection of the user's speech.

[0121] The invention may use an analog or digital filter 132 between the array microphone or conventional microphone 134 and the speech coder 138. The pass band of the filter can be set to optimize the signal to noise ratio at the input to the speech recognition engine 120. In some embodiments, the filter is adaptive, using band shaping combined with notch filtering to reject narrow-band noise. One embodiment employs adaptive echo cancellation in the filter. The echo cancellation helps prevent cross talk between output from the text to speech engine and detection of the user's speech as well as suppression of environmentally caused echoes.

Google Exhibit 1019
Page 00542

Algorithms comparing the background noise to the signal received from the users speech may be used to optimize the band-shaping parameters of the adaptive filter.

[0122] The speech received by the array microphone 134 and passed through the filter 132 may be sent to the speech digitizer or coder 138. The speech coder may use adaptive lossy audio compression to optimize bandwidth requirements for the transmission of the coded speech to the speech recognition engine 120 over a wireless link. The lossy coding is optimized to preserve only the components of the speech signal required for optimal recognition. Further, the lossy compression algorithms that may be used are designed to prevent even momentary gaps in the signal stream, which can cause errors in the speech recognition engine. The digitized speech may be buffered in the coder and the coder may adapt the output data rate to optimize the use of the available bandwidth. The use of the adaptive speech coder is particularly advantageous when a band-limited wireless link is used between the coder and the speech recognition engine.

[0123] The microphone can be complemented with an analog or digital (i.e., Voice over IP) speech interface. This interface allows a remote user to connect to the system and interact with it in the same manner possible if they were physically present.

[0124] In an alternative embodiment, the array microphone can be replaced by a set of physically distributed microphones or headsets worn by the users. The distributed microphones can be placed in different parts of a vehicle, different parts of a room or in different rooms of a building. The distributed microphones can create a three-dimensional array to improve signal to noise ratio. The headset can use a wireless or wired connection.

[0125] While the invention is intended to be able to accept most any natural language question or command, ambiguity may be a problem. To assist users formulate concise questions and commands, the system can support a voice query language. The language may be structured to allow a variety of queries and commands with minimal ambiguity. Thus, the voice query language helps users clearly specify the keywords or contexts of the question or command along with the parameters or criteria. The language can provide a grammar to clearly specify the keyword used to determine the context and present a set of one or criteria or parameters. A user asking a question or

Google Exhibit 1019
Page 00543

stating a command in the voice query language may nearly always be guaranteed to receive a response.

[0126] The voice query language may be sensitive to the contents of the context stack, wherein a context defines a set of questions that can be activated or deactivated during a conversation. According to one embodiment, each agent may designate one context to be the root context that defines base algorithms that the agent implements. Thus, follow-on questions can be asked using an abbreviated grammar, since key words and criteria can be inherited from the context stack. For example, the user can simply ask about another keyword if the criteria of the question remain constant.

[0127] The system 90, may provide built in training capabilities to help users learn the best methods to formulate their questions and commands. The interactive training allows users to audibly or visibly see the machine interpretation of their queries and provides suggestions on how to better structure a query. Using the interactive training, users can quickly become comfortable with the voice query language and, at the same time, learn how to optimize the amount of information required with each step of a dialog.

[0128] The output of the speech coder 122 may be fed to the speech recognition engine 120. The speech recognition engine 120 recognizes words and phrases, using information in the context description grammar 112 and passes these to the parser 118 for interpretation. The speech recognition engine 120 may determine the user's identity by voice and name for each utterance. Recognized words and phrases may be tagged with this identity in all further processing. Thus, as multiple users engage in overlapping sessions, the tags added by the speech recognition engine 120 to each utterance allows other components of the system 90 to tie that utterance to the correct user and dialog. The user recognition capability may further be used as a security measure for applications, such as auctions or online shopping, where this is required. Voice characteristics of each user may be contained in the user profile 110.

[0129] According to one embodiment of the invention, users may start a dialog with the system 90 when they first address it. This can be done by speaking a generic word ("computer") or addressing a specific name ("Fred"), which may be generally tied to a system personality 108. Once the user starts the dialog, it may be recognized by the speech recognition engine 120, using unique characteristics of the user's speech. At

Google Exhibit 1019
Page 00544

the end of a dialog or to interrupt a dialog, the user may utter a dismissal word ("good bye").

[0130] According to another embodiment of the invention, the system 90 may employ a speech recognition engine 124 that gains improved word recognition accuracy using data from context description grammar 112, user profiles 110, and the agents 106, among other components. At the same time, the fuzzy set possibilities or prior probabilities for the words in the context description grammar may be dynamically updated to maximize the probability of correct recognition at each stage of the dialog. The probabilities or possibilities may be dynamically updated based on a number of criteria including the application domain, the questions or commands, contexts, the user profile and preferences, user dialog history, the recognizer dictionary and phrase tables, and word spellings, among other criteria.

[0131] For uncommon words or new vocabulary words, the user may be given the option to spell the words. The spelling can be done by saying the names or the letters or using a phonetic alphabet. The phonetic alphabet can be a default one or one of the user's choosing.

[0132] Alternatively, when a user submits a word that is not recognized at all or is not correctly recognized by the speech recognition engine 120, then the user may be asked to spell the word. The speech recognition engine 120 determines this condition based on confidence level for the scoring process. The word may be looked up in a dictionary and the pronunciation for the word is added to either the dictionary, the agent 106, or the user's profile 110. The word pronunciation can then be associated with the domain, the question, the context and the user. Through this process, the speech recognition engine learns with time and improves in accuracy. To assist users in spelling words, an individualized phonetic alphabet can be used. Each user can modify the standard phonetic alphabets with words, which they can remember more easily.

[0133] Once the words and phrases have been recognized by the speech recognition engine 120, the tokens and user identification may be passed to the parser 118. The parser examines the tokens for the questions or commands, context and criteria. The parser may determine a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents. The context of a question or command may determine the domain

Google Exhibit 1019
Page 00545

and, thereby, the domain agent 156, if any, to be invoked. For example, a question with the keywords "temperature" implies a context value of weather for the question. Within a different dialog, the keyword "temperature" can imply a context for a measurement. The parser dynamically receives keyword and associated prior probability or fuzzy possibility updates from the system agent 150 or an already active domain agent 156. Based on these probabilities or possibilities the possible contexts are scored and the top one or few are used for further processing.

[0134] The parser 118 uses a scoring system to determine the most likely context or domain for a user's question and/or command. The score is determined from weighing a number of factors including, the user profile 110, the domain agent's data content and previous context. Based on this scoring, the system 90 invokes the correct agent. If the confidence level of the score is not high enough to ensure a reliable response, the system 90 may ask the user to verify whether the question and/or command is correctly understood.

[0135] In general, the question that is asked by the system 90 may be phrased to indicate the context of the question including all criteria or parameters. For example, the question can be in the form of: "Did I understand that you want such-and-such" If the user confirms that the question is correct the system proceeds to produce a response. Otherwise, the user can rephrase the original question, perhaps adding additional information to remove ambiguity, or the system can ask one or more questions to attempt to resolve the ambiguity.

[0136] Once the context for the question or command has been determined, the parser 118 can invoke the correct agent 156, 150. To formulate a question or command in the regular grammar used by agents, the parser will preferably determine required and optional values for the criteria or parameters. These criteria may have been explicitly supplied by the user or may need to be inferred. The parser may make use of the criteria handlers 152 supplied by the system agent. The criteria handlers can provide context sensitive procedures for extracting the criteria or parameters from the user's question or command. Some criteria may be determined by executing algorithms in the agent, while others may be determined by applying probabilistic or fuzzy reasoning to tables of possible values. Prior probabilities or fuzzy possibilities and associated values may be received from a number of sources including, for example, the history of the dialog, the user profile 110, and the agent. Based on user

Google Exhibit 1019
Page 00546

responses, the prior probabilities or fuzzy possibilities may be updated as the system learns the desired behavior. For a weather context, examples of criteria include, location, date and time. Other criteria can include command criteria (i.e., yes/no, on/off, pause, stop), and spelling. Special criteria handlers are available from the system agent for processing lists, tables, barge-in commands, long strings of text and system commands.

[0137] The criteria handlers 152 can operate iteratively or recursively on the criteria extracted to eliminate ambiguity. This processing may help reduce the ambiguity in the user's question or command. For example, if the user has a place name (or other proper noun) in their utterance the parser 118 can use services of the domain agent 156 to look up tables in the databases 102 for place names or can attempt to determine which word is the proper noun from the syntax of the utterance. In another example, the user asks, "what about flight one hundred and twenty too?" The parser and domain agent use flight information in the database and network information along with context to determine the most plausible interpretation among; flight 100 and flight 20 also, flight 100 and flight 22, flight 122, etc.

[0138] Once the context and the criteria are determined, the parser 118 may form the question or command in a standard format or hierarchical data structure used for processing by the agents 150, 156. The parser 118 may fill in all required and some optional tokens for the grammar of the context. Often the tokens must be transformed to values and forms acceptable to the agents. The parser obtains the required transformations from the agents, dialog history or user profile 110. Examples of transformations or substitutions performed by the parser on tokens include: 1) substituting a stock symbol for a company name or abbreviation; 2) substituting a numerical value for a word or words; 3) adding a zip code to an address; and, 4) changing a place or other name to a commonly used standard abbreviation.

[0139] The agents 150, 156 may receive a command or question once the parser 118 has placed it in the required standard format. Based on the context, the parser can evoke the correct agent to process the question or command.

[0140] Commands can be directed to the system or to an external entity. System commands are generally directed to the system agent 150. Commands for external entities are generally processed by a domain agent 156, which includes the command context and behavior for the external entity.

Google Exhibit 1019
Page 00547

[0141] Specific questions may be generally directed to one of the domain agents 156. The real-time selection of the correct agent allows the invention to dynamically switch contexts. Based on the question, command or context and the parameters or criteria, the domain agent may create one or more queries to one or more local or external information sources. Questions can be objective or subjective in nature. Results for objective questions can often be obtained by structured queries to one or more local or network information sources. Even for objective questions, the system may need to apply probabilistic or fuzzy set analysis to deal with cases of conflicting information or incomplete information. Information to answer subjective questions is generally obtained by one or more ad-hoc queries to local or network data sources, followed by probabilistic or fuzzy set evaluation of the one results to determine a best answer.

[0142] Once the domain agent 156 has formulated the one or more queries, they may be sent to local and/or network information sources. The queries may be performed in an asynchronous manner to account for the fact that sources respond at different speeds or may fail to respond at all. Duplicate queries can be sent to different information sources to ensure that at least one source responds with a useful result in a timely manner. Further, if multiple results are received in a timely manner, they can be scored by the system to determine which data is most reliable or appropriate. Examples of data sources accommodated include, HTTP data sources, sources with meta-data in various formats including XML, measurement data from sensors using various formats, device 32 setting parameters, entertainment audio, video and game files including MP3, databases using query languages and structured responses such as SQL, and other data sources.

[0143] The local information sources can be stored in one or more system databases 102 or can be on any local data storage such as a set of CDs or DVDs in a player or other local data storage. In other cases, local information can be obtained from vehicle system settings or measurement devices. Network information sources can be connected to, the control and device interfaces 30, the data interfaces 26, the Internet 42 or other network and accessed through a series of plug-ins or adaptors, known a pluggable sources, in the network interface 116. The pluggable sources are capable of executing the protocols and interpreting the data formats for the data sources of interest. The pluggable source provides information scrapping forms and procedures

Google Exhibit 1019
Page 00548

for each source to the domain agents 156. If a new type of data source is to be used a new plug-in or adaptor can be added to the appropriate interface.

[0144] The domain agent 156 can evaluate the results of the one or more queries as they arrive. The domain agent may score the relevance of the results based on results already received, the context, the criteria, the history of the dialog, the user profile 110 and domain specific information using probabilistic or fuzzy scoring techniques. Part of the dialog history is maintained in a context stack. The weight of each context for the scoring may be based on the relevance of one context to another and the age of the contexts. Other scoring variables can be associated through the context stack. Contexts can also be exclusive, so that previous contexts have no weight in the scoring.

[0145] Based on the on-going scoring processes, the domain agent 156 may determine if a single best answer can be extracted. For most questions, the desired result may include a set of tokens that may be found to formulate an answer. Once a value has been found for each of these tokens, the results are ready for presentation to the user. For example, for a question on weather, the tokens can include the date, day of week, predicted high temperature, predicted low temperature, chance of precipitation, expected cloud cover, expected type of precipitation and other tokens. Results processed in this manner may include error messages. For subjective questions, this determination is made by determining a most likely answer or answers, extracted by matching of the results received. If no satisfactory answer can be inferred from the results of the query, the agent can do one of the following:

1. Ask the user for more information, typically through the speech interface, and based on the results obtained formulate new queries. This approach is applied when an irresolvable ambiguity arises in the formulation of a response.
2. Formulate new queries based on the results received from the first set of queries. This approach is typically applied in cases where the responses received do not include all the required information. Information sources to queries can be inferred from the results already obtained (i.e., links in an HTML document or measurements or settings from other devices 32) or from other sources. Using this approach one or more sets of queries and responses can be chained without the need for action by the user.

Google Exhibit 1019
Page 00549

3. Wait for additional queries to return results.

[0146] In any case, the domain agent 156 may continue to make queries and evaluate results until a satisfactory response is constructed. In doing so, the agent can start several overlapping query paths or threads of inquiry, typically mediated by the event manager 100. This technique, combined with the use of asynchronous queries from multiple data sources provides the real-time response performance required for a natural interaction with the user.

[0147] The domain agent 156 may apply conditional scraping operations to each query response as it is received. The conditional scraping actions may depend on the context, the criteria, user profile 110, and domain agent coding and data. For each token to be extracted a scraping criteria 152 can be created using the services of the system agent 150. The scraping criteria may use format specific scraping methods including, tables, lists, text, and other methods. One or more scraping criteria can be applied to a page or results set. Once additional results are received, the domain agent can create new scraping criteria to apply to results already acquired. The conditional scarping process removes extraneous information, such as graphics, which need not be further processed or stored, improving system performance.

[0148] Specific commands are generally directed to one of the domain agents 156. The real-time selection of the correct agent allows the invention to dynamically switch contexts. Command oriented domain agents 156 evaluate the command and the state of vehicle systems, system capabilities, and measurements to determine if the command can be executed at all or if doing so will exceed operating or safety limits. If the command is ambiguous or cannot be executed for some other reason, the system may ask the user for more information or may suggest what the problem is and a likely approach to the solution. The domain agent may format the command for the specific device 32 and control and device interface 30. This formatting may involve variable substitution, inference of missing values and other formatting. Variable substitution and inference depends on the command context, the user profile 110, command history, state of vehicle systems and measured values, and other factors. A complex command can result in more atomic commands being sent to multiple devices, perhaps in a sequence. The sequence and nature of subsequent commands may depend on the previous commands, results of pervious commands, device settings and other measurements. As a command is executed, measurements are made

Google Exhibit 1019
Page 00550

and results collected to determine if the execution was correct and the desired state or states were reached.

[0149] Once the domain agent 156 has created a satisfactory response to a question, or to a command, the agent may format that response for presentation. Typically, the domain agent can format the response into the markup format used by the text to speech engine 124. The domain agent may format the result presentation using available format templates and based on the context, the criteria, and the user profile 110. The domain agent may perform variable substitutions and transformations to produce a response best understood and most natural to the user. The domain agent may vary the order of presentation of tokens and the exact terminology used to create a more natural response to the user. The domain agent may also select the presentation personality 108 to be used.

[0150] For both command and query responses, the domain agent 156 may select the presentation template, determine order of presentation for tokens and determine variable substitutions and transformations using probabilistic or fuzzy set decision methods. The template used to form the presentation can be from the domain agent itself or from the user profile 110. The user profile can completely specify the presentation format or can be used to select and then modify an existing presentation format. Selection and formatting of presentation template can also depend on the presentation personality 108. At the same time, the characteristics of the personality used for the response are dynamically determined using probabilities or fuzzy possibilities derived from the context, the criteria, the domain agent itself and the user profile 110.

[0151] The domain agent 156 may apply a number of transformations to the tokens before presentation to the user. These variable substitutions and transformations may be derived from a number of sources including, domain information carried by the agent, the context, the token values, the criteria, the personality 108 to be used, and the user profile 110. Examples of variable substitutions and transformations include: 1) substitution of words for numbers; 2) substitution of names for acronyms or symbols (i.e., trading symbols); 3) use of formatting information derived from the information sources (i.e., HTML tags); 4) nature of the response including, text, long text, list, table; 5) possible missing information or errors; 6) units for measurement

Google Exhibit 1019
Page 00551

(i.e., English or metric); and, 7) preferred terminology from the user profile or presentation personality 108.

[0152] The invention may provide special purpose presentation capabilities for long text strings, tables, lists and other large results sets. Domain agents 156 may use special formatting templates for such results. The system agent 150 can provide special criteria handlers 152 for presentation and user commands for large results sets. The presentation templates used by the domain agents for large results sets typically include methods for summarizing the results and then allowing the user to query the result in more detail. For example, initially only short summaries, such as headlines or key numbers, are presented. The user can then query the results set further. The criteria handlers provide users with the capability to browse large results sets. Commands provided by the criteria handlers for large results sets include, stop, pause, skip, rewind, start, and forward.

[0153] Some information, in formats such as video, pictures and graphics, may be best presented in a displayed format. The domain agents 156 apply suitable presentation templates in these cases and present the information through the non-speech interface 114. The system agent 150 provides special criteria handlers 152 for presentation and user commands for display presentation and control.

[0154] Although particular embodiments of the invention have been shown and described, it will be understood that it is not intended to limit the invention to the embodiments that are disclosed and it will be obvious to those skilled in the art that various changes and modifications may be made without departing from the spirit and scope of the invention. Thus, the invention is intended to cover alternatives, modifications, and equivalents, which may be included within the spirit and scope of the invention as defined by the claims.

Google Exhibit 1019
Page 00552

We claim:

1. A mobile device for processing speech and non-speech communications, comprising:

an identification module that identifies a user that communicates with the mobile device;

a matching module that matches the identified user to one or more profiles;

a transcription module that transcribes the speech and the non-speech communications to create a speech-based textual message and a non-speech-based textual message, wherein the transcription module accesses the one or more profiles;

a merging module that merges the speech-based textual message and the non-speech-based textual message to generate a query;

a search module that searches the query for text combinations;

a comparison module that compares the text combinations to entries in a context description grammar;

a plurality of domain agents that are associated with the context description grammar;

a scoring module that provides relevance scores based on results from the comparison module;

a domain agent selector that selects domain agents based on results from the scoring module; and

a response generating module that communicates with the selected domain agents to obtain content that is gathered by the selected domain agents and that generates a response from the content, wherein the content is arranged in a selected order based on results from the scoring module.

2. The mobile device according to claim 1, wherein the response generating module generates an aggregate response that includes the content gathered by the selected domain agents.

3. The mobile device according to claim 1, wherein the identification module supports multiple users that access the mobile device during a same session.

Google Exhibit 1019
Page 00553

4. The mobile device according to claim 3, wherein the session may include an overlapping session or an interleaved session.

5. The mobile device according to claim 3, wherein the identification module verifies an identify of the multiple users based on voiceprint matching or password matching.

6. The mobile device according to claim 4, wherein, during the interleaved session, the mobile device processes one or more queries in an order of receipt.

7. The mobile device according to claim 1, wherein, during the interleaved session, the mobile device processes one or more queries based on an order determined by a length of the queries.

8. The mobile device according to claim 1, wherein the mobile device includes (i) a personal digital assistant, (ii) a cellular telephone, (iii) a portable computer, (iv) a desktop computer, or any combination of (i) to (iv).

9. The mobile device according to claim 1, wherein the mobile device receives follow-up speech and non-speech communications and wherein the transcription module transcribes the follow-up speech and non-speech communications to create a follow-up speech-based textual message and a follow-up non-speech-based textual message.

10. The mobile device according to claim 9, wherein the merging module merges the follow-up speech-based textual message and the follow-up non-speech-based textual message to generate a follow-up query.

11. The mobile device according to claim 1, further comprising a personality module that facilitates formatting the response.

12. The mobile device according to claim 1, wherein the mobile device is configured to communicate with a context stack that includes one or more contexts that are selected based on the query.

13. The mobile device according to claim 12, wherein the scoring module determines the one or more contexts based on at least applying prior probabilities or fuzzy possibilities to (i) keyword matching, (ii) user profiles, (iii) a dialog history, or any combination of (i) to (iii).

Google Exhibit 1019
Page 00554

14. The mobile device according to claim 1, wherein at least one of the domain agents creates and directs a request to a local information source or a network information source.

15. The mobile device according to claim 1, wherein at least one of the domain agents creates and directs a command to a remote or local device.

16. The mobile device according to claim 10, wherein the follow-up query is associated with a same context as the query.

17. The mobile device according to claim 1, wherein the mobile device supports Internet browsing and wherein the generated query is used to conduct the Internet browsing.

18. The mobile device according to claim 1, wherein the mobile device supports multimodal communications.

19. The mobile device, comprising:

an object that is accessible by the mobile device;

an input device that receives speech and non-speech annotations corresponding to the object;

a transcription module that transcribes the speech and the non-speech annotations to create a speech-based textual message and a non-speech-based textual message; and

an association module that associates the object with the speech-based textual message and the non-speech-based textual message.

20. The mobile device according to claim 19, further comprising:

a classification module that classifies the speech-based textual message and the non-speech-based textual message; and

a searching module that searches the speech-based textual message and the non-speech-based textual message based on a search request.

21. The mobile device according to claim 19, wherein the association module associates the object with the speech and non-speech annotations.

22. The mobile device according to claim 21, further comprising:

a classification module that classifies the speech and non-speech annotations; and

Google Exhibit 1019
Page 00555

a searching module that searches the speech and non-speech annotations based on a search request.

23. The mobile device according to claim 19, wherein the object includes digital media.

24. The mobile device according to claim 23, wherein the object includes a digital photograph, a calendar entry, an email message, an instant message, a phonebook entry, a voicemail entry or a digital movie.

25. The mobile device according to claim 19, wherein the association module associates the object to the speech and non-speech annotations, the speech-based textual message, and the non-speech-based textual message.

26. The mobile device according to claim 25, wherein the object is retrievable using (i) a voice command, (i) a text command, or a combination of (i) and (ii).

27. The mobile device according to claim 19, wherein the object is stored (i) in the mobile device, (ii) at a server, or a combination of (i) and (ii).

28. The mobile device according to claim 19, wherein the object includes metadata.

29. The mobile device according to claim 28, wherein the metadata includes global positioning system information, environmental information, or geographic information.

30. The mobile device according to claim 28, further comprising a metadata searching module that enables searching of the metadata.

31. A system, comprising:

a plurality of mobile devices that support multimodal communications, wherein two or more mobile devices are associated with a user; and

a context module that is coupled to the plurality of mobile devices, wherein a same context is maintained at the two or more mobile devices that are associated with the user.

32. The system according to claim 31, wherein the plurality of mobile devices comprise:

an identification module that identifies a user that communicates with the mobile device;

Google Exhibit 1019
Page 00556

a matching module that matches the identified user to one or more profiles;

a transcription module that transcribes the speech and the non-speech communications to create a speech-based textual message and a non-speech-based textual message, wherein the transcription module accesses the one or more profiles;

a merging module that merges the speech-based textual message and the non-speech-based textual message to generate a query;

a search module that searches the query for text combinations;

a comparison module that compares the text combinations to entries in a context description grammar;

a plurality of domain agents that are associated with the context description grammar;

a scoring module that provides relevance scores based on results from the comparison module;

a domain agent selector that selects domain agents based on results from the scoring module; and

a response generating module that communicates with the selected domain agents to obtain content that is gathered by the selected domain agents and that generates a response from the content, wherein the content is arranged in a selected order based on results from the scoring module.

33. The system according to claim 31, wherein the context module is configured to broadcast the same information to two or more mobile devices and to receive a response from one of the mobile devices.

34. The system according to claim 31, wherein the context module broadcasts a first type of information and receives a second type of information.

35. The system according to claim 34, wherein first type of information and the second type of information include speech information or non-speech information.

36. The system according to claim 35, wherein the first type of information includes non-speech information that is presented on a graphical user interface and the second type of information includes speech information that is applied to search a network.

Google Exhibit 1019
Page 00557

37. The system according to claim 36, wherein the first type of information and the second type of information correspond to the same context.

38. A method of processing speech and non-speech communications in a multimodal environment, comprising:

receiving the speech and non-speech communications;

identifying a user that provides the speech and non-speech communications;

matching the identified user to one or more profiles;

transcribing the speech and non-speech communications to create a speech-based textual message and a non-speech-based textual message, wherein the transcribing includes accessing the one or more profiles;

merging the speech-based textual message and the non-speech-based textual message to generate a query;

searching the query for text combinations;

comparing the text combinations to entries in a context description grammar;

accessing a plurality of domain agents that are associated with the context description grammar;

generating a relevance score based on results from comparing the text combinations to entries in the context description grammar;

selecting one or more domain agents based on results from the relevance score;

obtaining content that is gathered by the selected domain agents; and

generating a response from the content, wherein the content is arranged in a selected order based on results from the relevance score.

39. The method according to claim 38, further comprising generating an aggregate response that includes the content that is gathered by the selected domain agents.

40. The method according to claim 38, further comprising:

receiving a follow-up speech and non-speech communications;

transcribing the follow-up speech and non-speech communications to create a follow-up speech-based textual message and a follow-up non-speech-based textual message; and

Google Exhibit 1019
Page 00558

merging the follow-up speech-based textual message and the follow-up non-speech-based textual message to generate a follow-up query.

41. The method according to claim 38, further comprising a personality module that communicates the response to a user.

42. The method according to claim 38, wherein the system supports multiple users during a same session.

43. The method according to claim 42, wherein the session may include an overlapping session or an interleaved session.

44. The method according to claim 42, wherein the multiple users are identified based on voiceprint matching or password matching.

45. The method according to claim 43, further comprising processing one or more queries in an order of receipt.

46. The method according to claim 43, further comprising processing one or more queries based on an order determined by a length of the queries.

47. The method according to claim 38, further comprising generating a context stack that includes one or more contexts that are selected based on the query.

48. The method according to claim 47, wherein the one or more contexts are generated based on applying prior probabilities or fuzzy possibilities to (i) keyword matching, (ii) user profiles, (iii) a dialog history, or any combination of (i) to (iii).

49. The method according to claim 38, further comprising conducting Internet browsing with the generated query.

Google Exhibit 1019
Page 00559

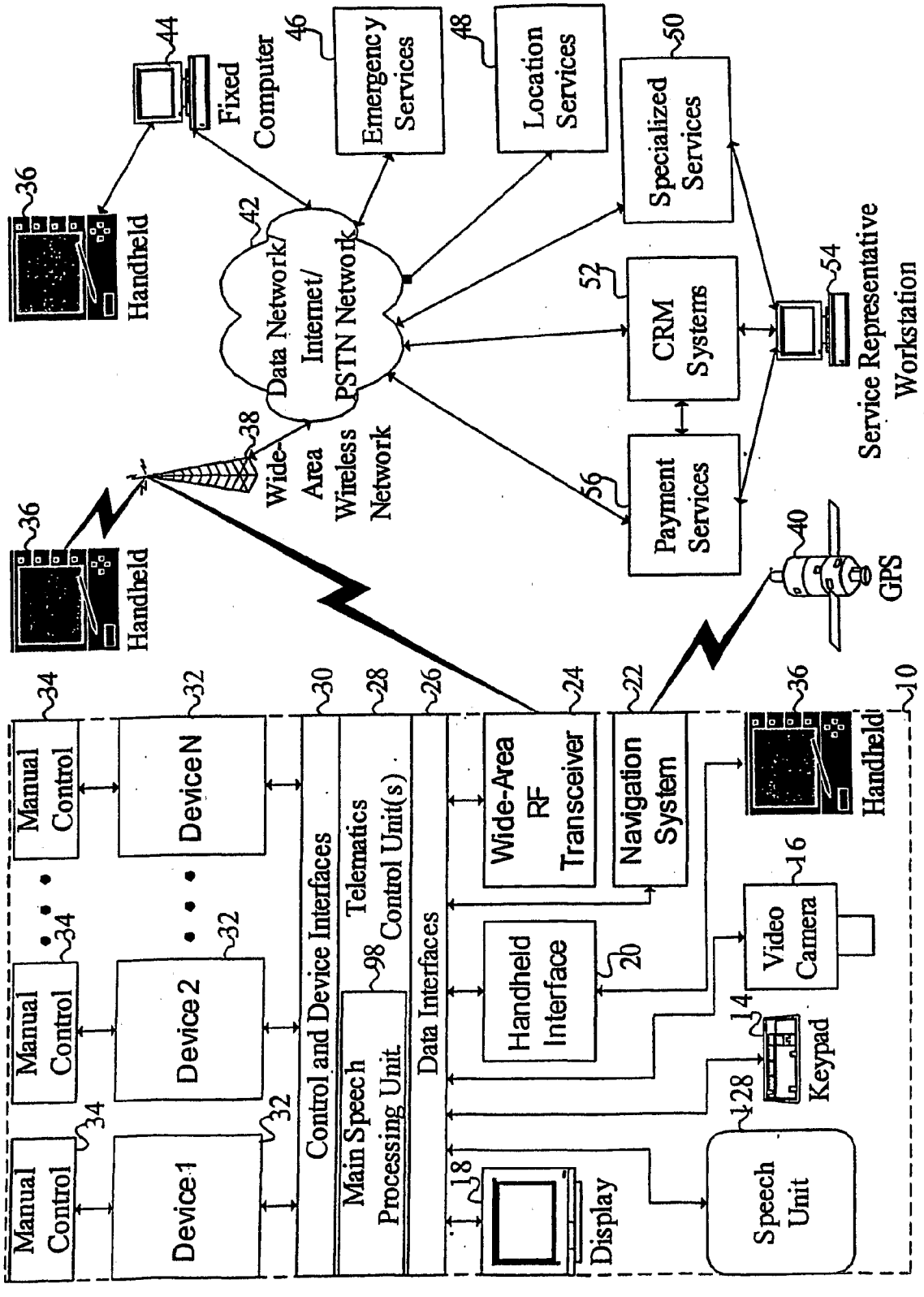

Handheld
36
Fixed Computer
44
Emergency Services
46
Location Services
48
Specialized Services
50
Data Network/ Internet/ PSTN Network
42
CRM Systems
52
Service Representative Workstation
54
Wide-Area Wireless Network
38
Payment Services
56
GPS
40
Manual Control
34
Device 1
32
Device 2
Device N
Control and Device Interfaces
30
Main Speech Processing Unit
Telematics Control Unit(s)
98
28
Data Interfaces
26
Handheld Interface
20
Wide-Area RF Transceiver
24
Navigation System
22
Video Camera
16
Keypad
14
Display
18
Speech Unit
128
Handheld
10

Figure 1

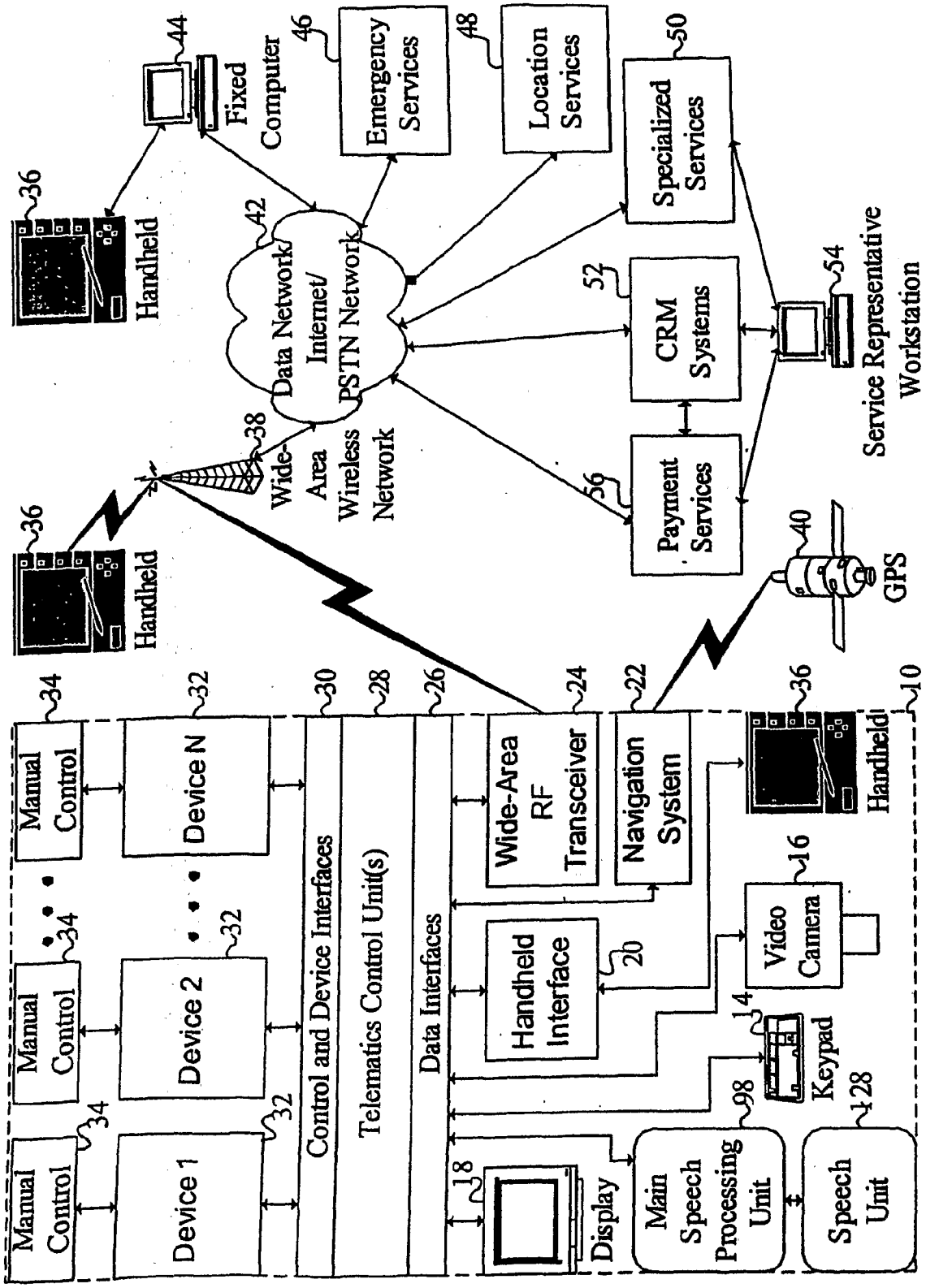

Manual Control
Device 1
Manual Control
Device 2
Manual Control
Device N
Control and Device Interfaces
Telematics Control Unit(s)
Data Interfaces
Display
Main Speech Processing Unit
Speech Unit
Keypad
Handheld Interface
Wide-Area RF Transceiver
Navigation System
Video Camera
Handheld
GPS
Payment Services
CRM Systems
Service Representative Workstation
Specialized Services
Location Services
Emergency Services
Fixed Computer
Data Network/ Internet/ PSTN Network
Wide-Area Wireless Network
Handheld
Handheld
10
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
98

Figure 2

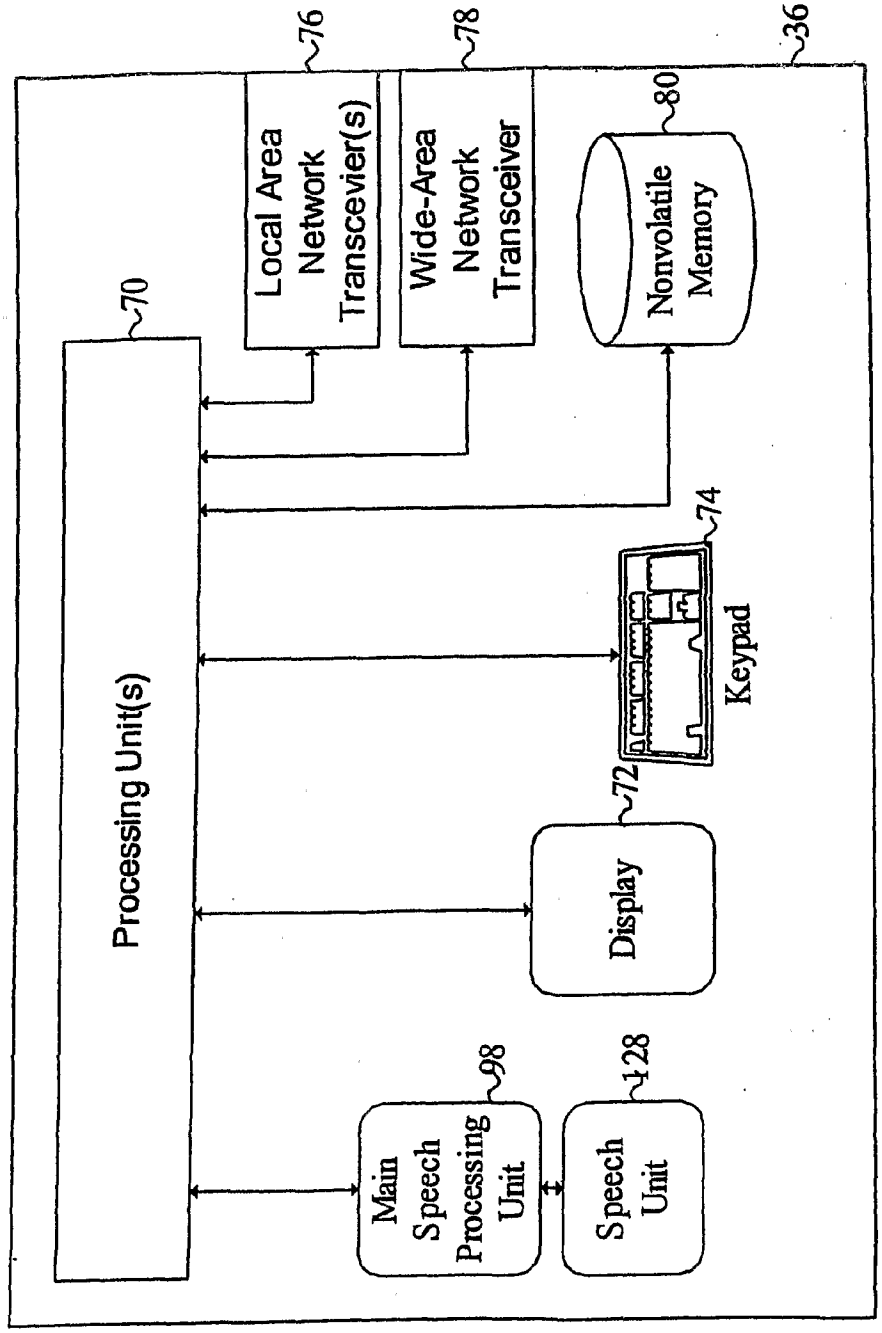

Processing Unit(s)
70
Local Area Network Transcevier(s)
76
Wide-Area Network Transceiver
78
Nonvolatile Memory
80
36
Keypad
74
Display
72
Main Speech Processing Unit
98
Speech Unit
128

Figure 3.

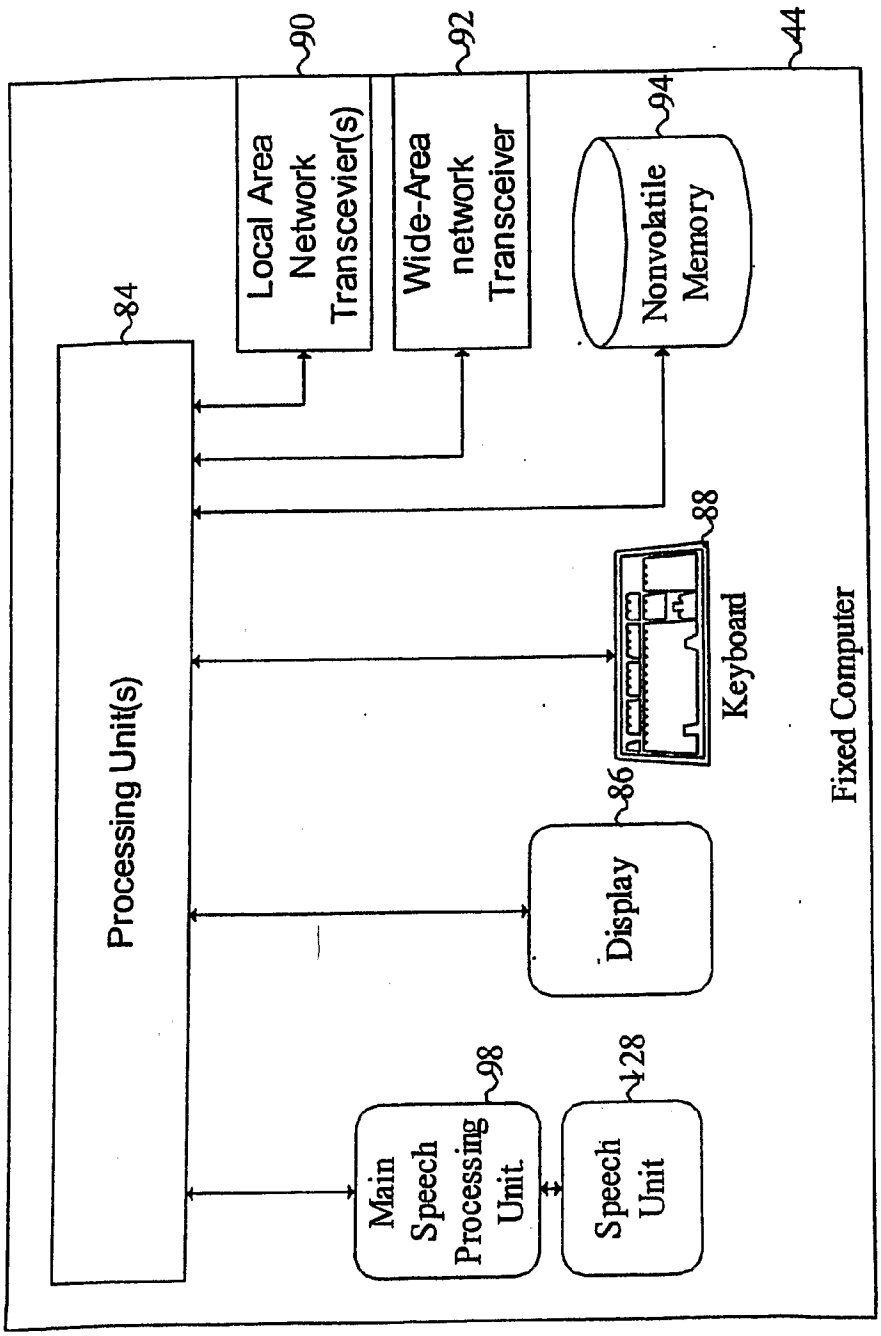

Fixed Computer
44
Processing Unit(s)
84
Local Area Network Transcevier(s)
90
Wide-Area network Transceiver
92
Nonvolatile Memory
94
Keyboard
88
Display
86
Main Speech Processing Unit
98
Speech Unit
128

Figure 4

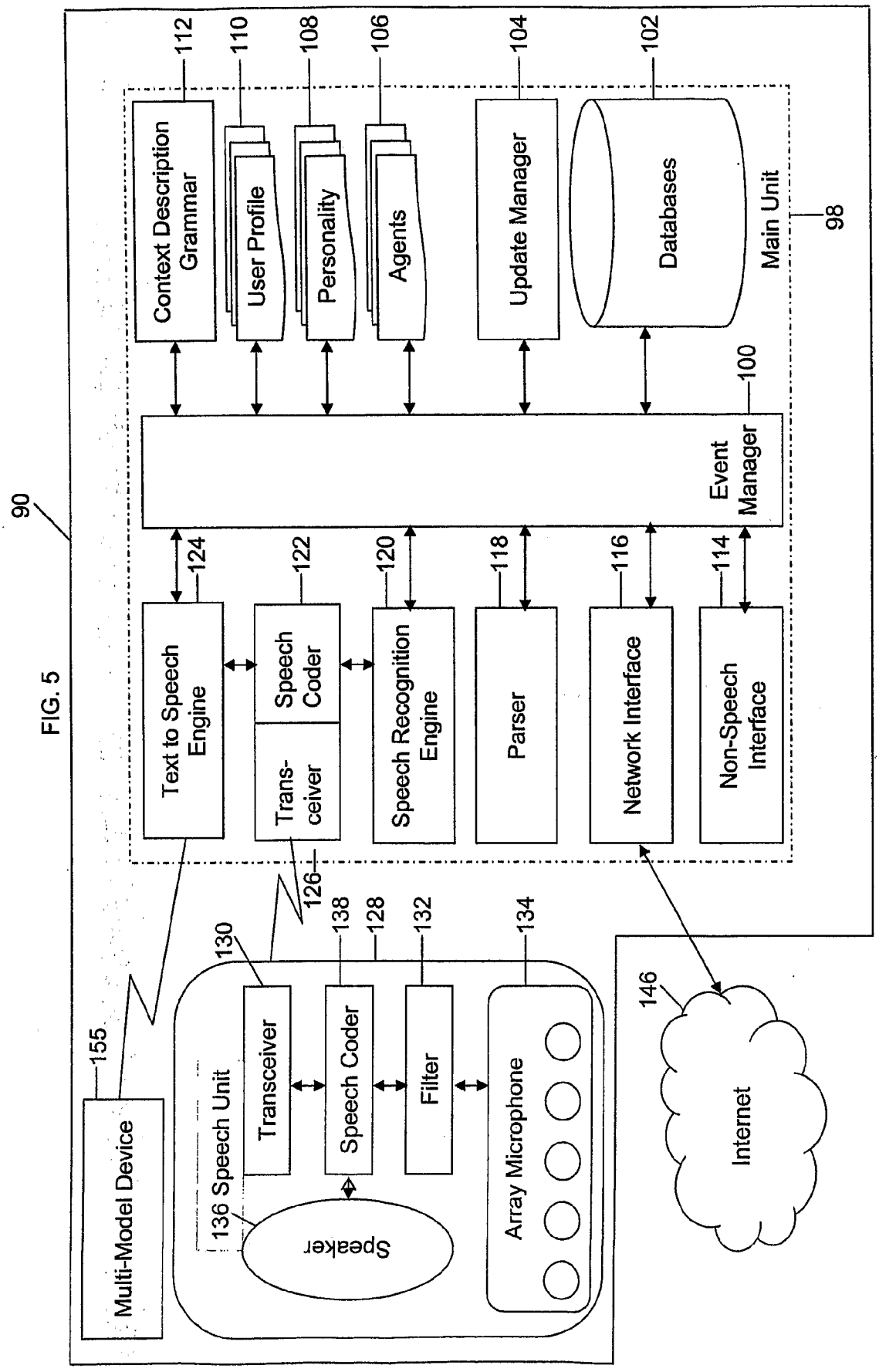

90
Main Unit
98
Context Description Grammar
112
User Profile
110
Personality
108
Agents
106
Update Manager
104
Databases
102
Event Manager
100
Text to Speech Engine
124
Speech Coder
Trans-ceiver
122
Speech Recognition Engine
120
Parser
118
Network Interface
116
Non-Speech Interface
114
126
130
138
128
132
134
Multi-Model Device
155
136 Speech Unit
Transceiver
Speech Coder
Filter
Array Microphone
Speaker
Internet
146

FIG. 5

FIG. 6

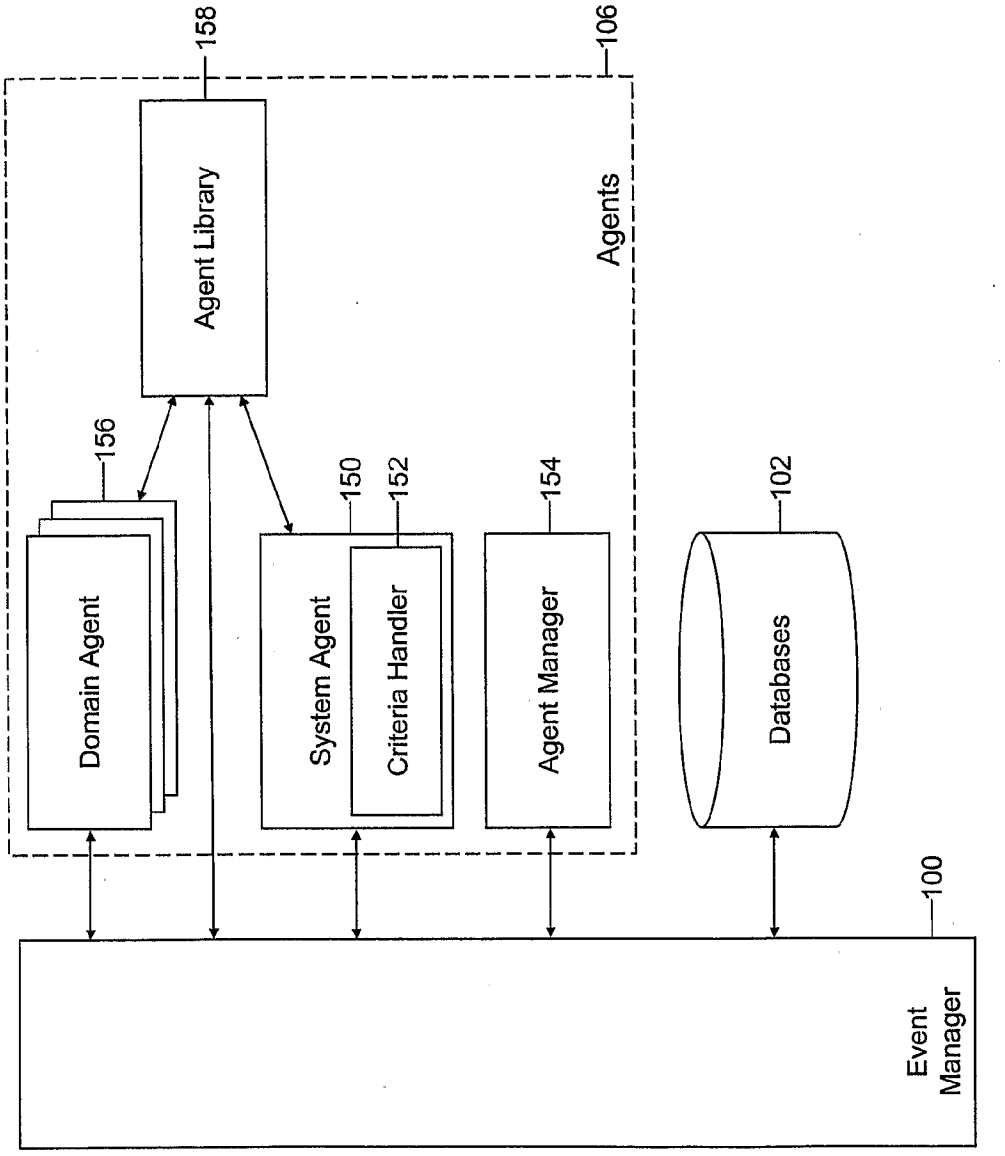
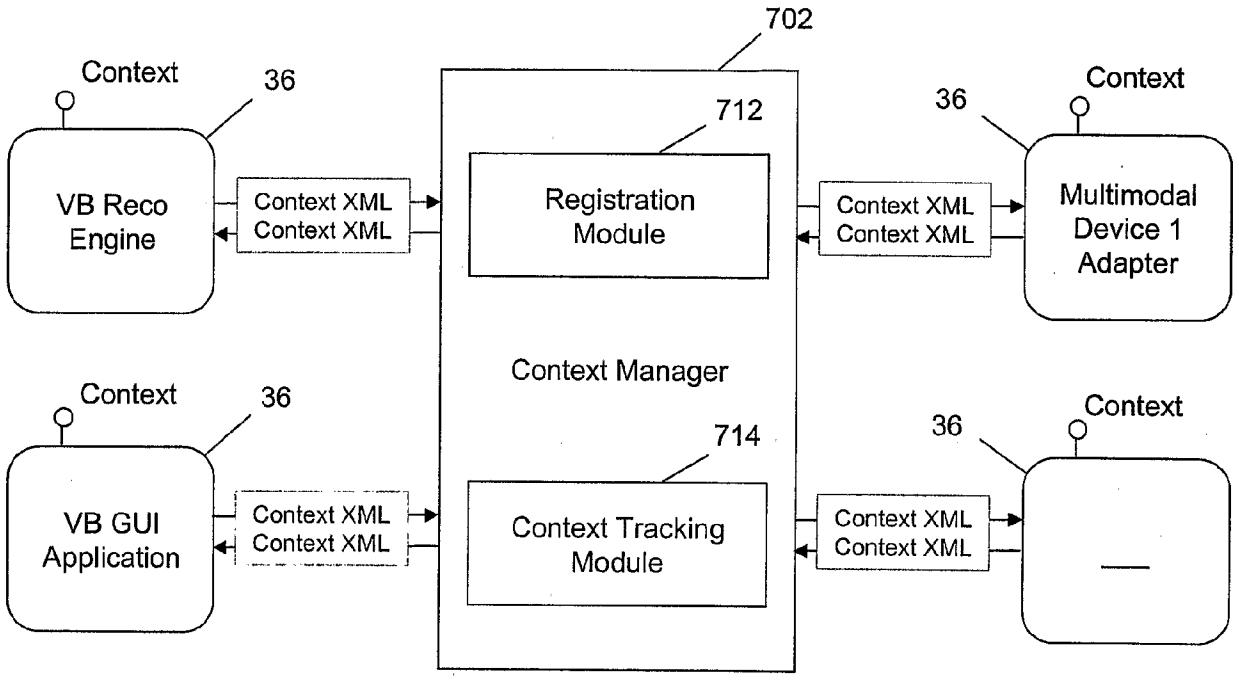

158
106
Agent Library
Agents
156
150
152
154
102
Domain Agent
System Agent
Criteria Handler
Agent Manager
Databases
100
Event Manager

FIG. 7


700
702
Context
36
VB Reco Engine
Context XML
Context XML
712
Registration Module
Context XML
Context XML
36
Context
Multimodal Device 1 Adapter
Context Manager
Context
36
VB GUI Application
Context XML
Context XML
714
Context Tracking Module
Context XML
Context XML
36
Context

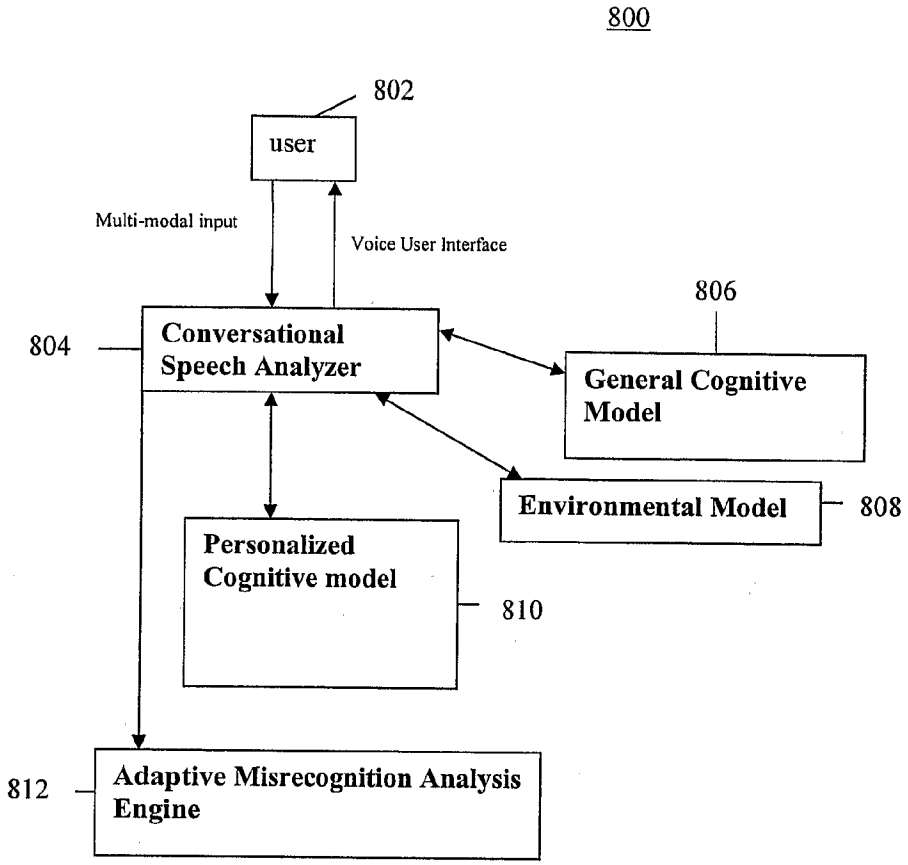

800
802
user
Multi-modal input
Voice User Interface
806
804
Conversational Speech Analyzer
General Cognitive Model
Environmental Model
808
Personalized Cognitive model
810
812
Adaptive Misrecognition Analysis Engine

FIGURE 8

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 March 2007 (08.03.2007)

PCT

(10) International Publication Number
WO 2007/027989 A2

(51) International Patent Classification:
G10L 15/18 (2006.01)

(21) International Application Number:
PCT/US2006/034184

(22) International Filing Date: 31 August 2006 (31.08.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/712,412 31 August 2005 (31.08.2005) US

(71) Applicant *(for all designated States except US)*: **VOICEBOX TECHNOLOGIES, INC.** [US/US]; 10505 NE 38th Place, Building 9, Kirkland, WA 98033 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DI CRISTO, Philippe** [FR/US]; 177 107th Avenue, N.E., #1007, Bellevue, WA 98004 (US). **KE, Min** [US/US]; 3221 201st Place, S.E., Bothell, WA 98012 (US). **KENNEWICK, Robert, A.** [US/US]; 2204 S. Eddy St., Seattle, WA 98108 (US). **TJALVE, Michael** [DK/US]; 14904 SE 47th Court, Bellevue, WA 98006 (US).

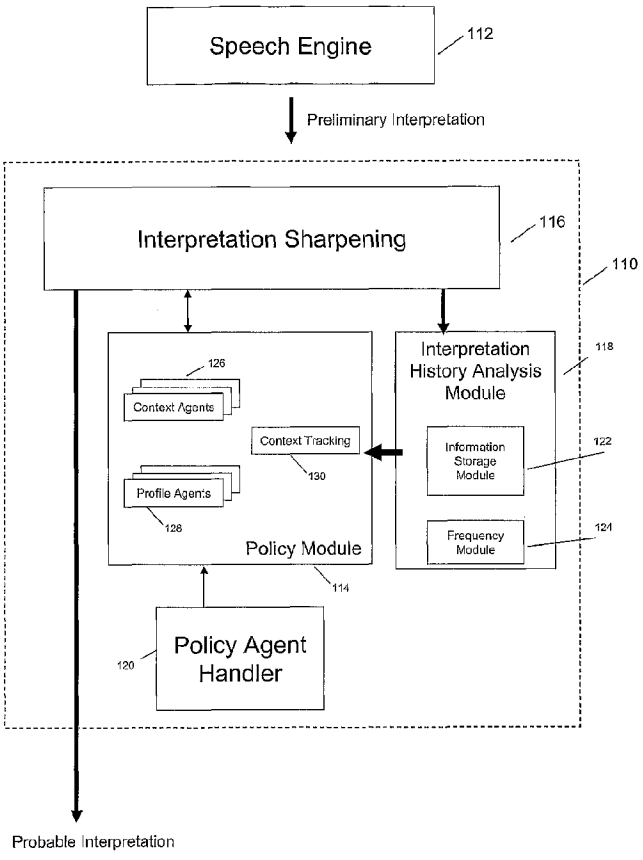
(74) Agents: **GATTO, James, G.** et al.; PILLSBURY WINTHROP SHAW PITTMAN LLP, P.O. Box 10500, Mclean, VA 22102 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

[Continued on next page]

(54) Title: DYNAMIC SPEECH SHARPENING



(57) Abstract: An enhanced system for speech interpretation is provided. The system may include receiving a user verbalization and generating one or more preliminary interpretations of the verbalization by identifying one or more phonemes in the verbalization. An acoustic grammar may be used to map the phonemes to syllables or words, and the acoustic grammar may include one or more linking elements to reduce a search space associated with the grammar. The preliminary interpretations may be subject to various post-processing techniques to sharpen accuracy of the preliminary interpretation. A heuristic model may assign weights to various parameters based on a context, a user profile, or other domain knowledge. A probable interpretation may be identified based on a confidence score for each of a set of candidate interpretations generated by the heuristic model. The model may be augmented or updated based on various information associated with the interpretation of the verbalization.

WO 2007/027989 A2

FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DYNAMIC SPEECH SHARPENING

CROSS REFERENCE TO RELATED APPLICATIONS

[001] This application claims the benefit of provisional U.S. Patent Application Serial No. 60/712,412, entitled "Dynamic Speech Sharpening," filed August 31, 2005, and which is hereby incorporated by reference in its entirety.

[002] This application is related to the following co-pending applications, which are hereby incorporated by reference in their entirety: "Systems and Methods for Responding to Natural Language Speech Utterance," Ser. No. 10/452,147, filed June 3, 2003, and published as US Patent Application Publication No. US 2004/0044516; "Mobile Systems and Methods for Responding to Natural Language Speech Utterance," Ser. No. 10/618,633, filed July 15, 2003, and published as US Patent Application Publication No. US 2004/0193420; "Systems and Methods for Responding to Natural Language Speech Utterance," Ser. No. 11/197,504, filed August 5, 2005; "System and Method of Supporting Adaptive Misrecognition in Conversational Speech," Ser. No. 11/200,164, filed August 10, 2005; and "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," Ser. No. 11/212,693, filed August 29, 2005.

FIELD OF THE INVENTION

[003] The invention is related generally to automated speech interpretation, and in particular, to enhancing the accuracy and performance of speech interpretation engines.

BACKGROUND OF THE INVENTION

[004] The field of automated speech interpretation is in increasingly higher demand. One use of automated speech interpretation is to provide voice requests to electronic devices. This may enable a user to simply speak to an electronic device rather than manually inputting requests, or other information, through pressing buttons, uploading information, or by other request input methods. Controlling various electronic devices through speech may enable the user to use the electronic devices more efficiently.

[005] However, existing technology in the field of automated speech interpretation, such as standard speech engines, automatic speech recognition (ASR), and other systems for interpreting speech, are unable to process a speech signal in an efficient manner, often constructing large grammars that include a large number of items,

Google Exhibit 1019
Page 00570

nodes, and transitions, which is a concern particularly for large-list recognition for embedded applications. If the grammar for an embedded application grows too much, it may not fit within the constrained space of an embedded application. With limited CPU power, response time and performance is easily affected due to the significant time needed to compile and load the grammar. Response time is further degraded because the speech engine has to parse through a large number of transition states to come up with a recognition result. Even when the speech engine is able recognize a word, the results are often unreliable because large grammars introduce greater risk of confusion between items as the size of the grammar increases. Existing techniques focus on reducing the size of a grammar tree by removing command variants or criteria items, but this approach strips functionality from the application.

[006] In addition to the performance problems associated with speech recognition engines that employ large word grammars, existing speech processing engines are unable to interpret natural human speech with a suitable accuracy to sufficiently control some electronic devices. In particular, speech interpretation engines still have substantial problems with accuracy and interpreting words that are not defined in a predetermined vocabulary or grammar context. Poor quality microphones, extraneous noises, unclear or grammatically incorrect speech by the user, or an accent of the user may also cause shortcomings in accuracy, such as when a particular sound cannot be mapped to a word in the grammar.

[007] In light of these and other problems, there is a need for enhanced automated speech interpretation that may interpret natural human speech with an augmented accuracy.

SUMMARY OF THE INVENTION

[008] According to one aspect of the invention, a system for enhancing automated speech interpretation is provided. The system may include a set of techniques for use in a speech-to-text engine to enhance accuracy and performance, for example, by reducing the search space of the speech engine. The problems with large-list recognition for embedded applications may also be improved by using phonetic dictation, which may recognize a phoneme string by disregarding the notion of words. The system may also use one or more post-processing techniques to sharpen an output of a preliminary speech interpretation made by a speech engine. The system may be modeled at least partially after one or more speech pattern

Google Exhibit 1019
Page 00571

recognition techniques used by humans, such as interpreting speech using words, word sequences, word combinations, word positions, context, phonetic similarities between two or more words, parts of speech, or other techniques.

[009] In one implementation of the invention, the system may receive a verbalization made by a user, where a speech engine may receive the verbalization. The speech engine may output information relating to a plurality of preliminary interpretations of the verbalization, where the plurality of preliminary interpretations represent a set of best guesses at the user verbalization. According to one aspect of the invention, the performance of the speech engine may be improved by using phoneme recognition. Phoneme recognition may disregard the notion of words, instead interpreting a verbalization as a series of phonemes, which may provide out-of-vocabulary (OOV) capabilities, such as when a user misspeaks or an electronic capture devices drops part of a speech signal, or for large-list applications, such as city and street names or song titles, for example. Phoneme recognition may be based on any suitable acoustic grammar that maps a speech signal into a phonemic representation. For example, the English language may be broken down into a detailed grammar of the phonotactic rules of the English language. Portions of a word may be represented by a syllable, which may be further broken down into core components of an onset, a nucleus, and a coda, which may be further broken down into sub-categories. Various different acoustic grammars may be formed as trees with various branches representing many different syllables forming a speech signal.

[010] According to another aspect of the invention, the performance of the speech engine and the phonemic recognition may be improved by pruning the search space used by the speech engine using a common phonetic marker. In one implementation, the acoustic grammar may be represented entirely by a loop of phonemes. In another implementation, the speech engine may reduce the search space by reducing the number of transitions in a grammar tree, thereby speeding up the process of compiling, loading, and executing the speech engine. For example, the phoneme loop may include a linking element between transitions. This may reduce the number of grammar transitions, such that grammar paths merge after a first transition and diverge after the linking element. In one implementation of the invention, a common acoustic element that is part of a speech signal may be used as the linking element. In one implementation of the invention, the acoustic element may be one that is very likely to be triggered even if it is unpronounced. For

Google Exhibit 1019
Page 00572

example, a *schwa* in the English language may be used as the linking element because *schwa* represents an unstressed, central vowel that is likely to be spoken even if unintended. Those skilled in the art will appreciate that acoustic models for different languages may use other frequently elided phonemes as linking elements to reduce the search space used by the speech engine.

[011] The speech engine may generate a plurality of preliminary interpretations representing a set of best guesses at the user verbalization. The preliminary interpretations may be stored in a matrix, array, or another form, and may be provided to an interpretation sharpening module to determine a probable interpretation of a verbalization made by a user by applying heuristic policies against the preliminary interpretation to identify dominant words and/or phrases. According to various aspects of the invention, the interpretation sharpening module may include a policy module that may manage and/or provide one or more policies that enable the sharpening module to generate a plurality of probable interpretations of the verbalization made by the user. For example, according to one aspect of the invention, the plurality of preliminary interpretations may be applied against one or more policies to generate a set of hypotheses as to a candidate recognition. Each hypothesis may be reanalyzed to generate an interpretation score that may relate to a likelihood of the probable interpretation being a correct interpretation of the verbalization, and the preliminary interpretation corresponding to the highest (or lowest) interpretation score may then be designated as a probable interpretation of the verbalization. The designated probable interpretation may be stored and used for augmenting the policies to improve accuracy.

[012] According to one aspect of the invention, the policy module may include one or more agents that represent domains of knowledge. The agents may compete using a weighted model to revise a preliminary interpretation by determining context and intent. Relevant substitution of suspect words and phrases may be based on phonetic similarities or domain appropriateness. A domain agent may include one or more domain parameters for determining a probable interpretation from a preliminary interpretation. For example, domain parameters may include a policy vocabulary, a word position in the verbalization, a word combination, a sentence structure, or other parameters. A domain agent may include a parameter weighting scheme that may weight individual parameters according to one or more weighting factors, such as, a frequency of use, a difficulty to understand, or other factors.

Google Exhibit 1019
Page 00573

[013] According to one aspect of the invention, the domain agents may revise a preliminary interpretation into a probable interpretation using phonetic fuzzy matching (PFM). In one implementation of the invention, the speech engine may output a phoneme stream that is applied against a model of phoneme feature similarities, drawn from domain agents, to identify a closest phonetic match using a multi-pass method. Domain agents may be loaded and prioritized into an M-Tree, which accounts for the possibility of the speech engine dropping or adding phonemes. An M-Tree may be an index structure that resolves similarity queries between phonemes using a closest-distance metric based on relative weightings of phoneme misrecognition, phoneme addition, and phoneme deletion. The M-Tree may be updated using an adaptive misrecognition model. For example, information about a verbalization and its components, as well as a probability that the probable interpretation was correct, may be stored and used for adapting the policy module for the user.

[014] In one implementation of the invention, the domain agents in the policy module may include one or more profile agents that may manage and/or provide one or more profile policies for revising a preliminary interpretation of a phoneme stream. For example, a profile agent may correspond to a user and may include one or more profile parameters tailored to the user. The profile agent may be used as a base policy to interpret any verbalizations made by the user. In other implementations, a profile agent may correspond to a particular language, a regional accent, or other profiles for interpreting a user verbalization. The profile agents may be augmented to enable the system to provide more accurate interpretations of verbalizations made by the user. The augmentation may include a user augmentation, such as providing additional vocabulary (e.g., names in an address book), one or more personalized pronunciations or other pronunciation information, or other user provided augmentations. The augmentation may also include a non-user provided augmentation, such as updates generated by a third party (e.g., a commercial administration and/or maintenance entity), or other non-user provided augmentations. The augmentation may be automated, such as adjusting a profile parameter-weighting scheme through an adaptive misrecognition model, as discussed above.

[015] In another implementation of the invention, the domain agents in the policy module may include one or more context agents that may manage and/or provide

Google Exhibit 1019
Page 00574

one or more context policies for revising a preliminary interpretation of a phoneme stream. For example, a context agent may correspond to a context, such as song titles, city and street names, movie titles, finance, or other contexts. A context agent may include one or more context parameters that may be tailored to a verbalization context. The context policy may enhance an ability of the system related to interpreting verbalizations made by the user in the verbalization context corresponding to the context agent. The context agents may be augmented to enable the system to provide more accurate interpretations of verbalizations made in a verbalization context corresponding to the context agent. The augmentation may include a user provided augmentation, a non-user provided augmentation, an automated augmentation, or other augmentations. The augmentation may be automated, such as adjusting a profile parameter-weighting scheme through an adaptive misrecognition model, as discussed above.

[016] According to various implementations of the invention, the policy module may determine which profile agents and/or which context agents to use through a set of heuristics provided in a context-tracking module. In one implementation, the context-tracking module may use phonetic fuzzy matching to track a series of verbalizations by the user to identify a verbalization context. The context-tracking module may utilize one or more M-Trees to track the series of verbalizations and determine a closest phonetic match. The context-tracking module may track one or more past verbalization contexts for the series of verbalizations, one or more current verbalization contexts for the series of verbalizations, and/or make predictions regarding one or more future verbalization contexts for the series of verbalizations. The policy module may utilize information about the verbalization context of the series of verbalizations generated by the context tracking module to manage and/or provide one or more profile and/or context agents.

[017] According to one aspect of the invention, the system may include an interpretation history analysis module that may enable the system to augment one or more domain agents based on an analysis of past interpretations related to previously interpreted verbalizations. The augmentations enabled by the interpretation history analysis module may include a user augmentation, a third-party augmentation, an automated augmentation, or other augmentations. The interpretation history analysis module may include an information storage module that may store interpretation information related to past verbalizations, such as one

Google Exhibit 1019
Page 00575

or more preliminary interpretations associated with a past verbalization, one or more interpretation scores associated with a past verbalization, one or more probable interpretations associated with a past verbalization, whether or not a past verbalization was interpreted correctly, or other information. A frequency module may be included in the interpretation history module, and the frequency module may use some or all of the information stored in the information storage module to generate one or more frequencies related to one or more past verbalizations. For example, the frequency module may calculate a frequency of word usage, word combinations, phonetic homonyms, interpretation errors for a particular verbalization, or other frequencies.

[018] The Information generated and/or stored by the interpretation history analysis module may be used to augment the profile and/or context agents in order to enhance the accuracy of subsequent interpretations. In some implementations, an adaptive misrecognition model may use one or more generated frequencies to augment one or more agents. For example, one or more parameters and/or weighting schemes of an agent or phonetic model may be augmented based on a frequency generated by the interpretation history analysis module. Other augmentations using information stored and/or generated by the interpretation history analysis module may be made, and the system may include a policy agent handler that may augment, update, remove, and/or provide one or more domain agents to the system. A domain agent may comprise a profile or context agent, and the policy agent handler may be controlled, directly or indirectly by a third party (e.g. a commercial entity, etc.). The policy agent handler may augment, update, remove, and/or provide domain agents to the system as part of a commercial agreement, such as a licensing agreement, a subscription agreement, a maintenance agreement, or other agreements.

[019] Other objects and advantages of the invention will be apparent to those skilled in the art based on the following detailed description and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[020] Fig. 1 illustrates an exemplary system for enhancing automated speech interpretation according to one implementation of the invention.

[021] Fig. 2 illustrates an exemplary grammar tree for enhancing the performance of a speech engine according to one implementation of the invention.

Google Exhibit 1019
Page 00576

[022] Fig. 3 illustrates an exemplary flow chart of a method for enhancing automated speech interpretation according to one implementation of the invention.

DETAILED DESCRIPTION

[023] Fig. 1 illustrates an exemplary system for enhancing automated speech interpretation according to one implementation of the invention. A speech-to-text processing engine 112 may receive a user verbalization, and speech engine 112 may generate one or more preliminary interpretations of the user verbalization. The preliminary interpretations may represent a set of best guesses as to the user verbalization arranged in any predetermined form or data structure, such as an array, a matrix, or other forms. In one implementation of the invention, speech engine 112 may generate the preliminary interpretations by performing phonetic dictation to recognize a stream of phonemes, instead of a stream of words. Phonemic recognition provides several benefits, particularly in the embedded space, such as offering out-of-vocabulary (OOV) capabilities, improving processing performance by reducing the size of a grammar, and eliminating the need to train Statistic Language Models (SLMs). Those skilled in the art will recognize other advantages of phonemic recognition.

[024] Speech engine 112 may apply the phoneme stream against one or more acoustic grammars that reliably map a speech signal to a phonemic representation in order to generate the plurality of preliminary interpretations. Characteristics of a speech signal may be mapped to a phonemic representation to construct a suitable acoustic grammar, and various acoustic grammars may be included in speech engine 112 to generate one or more preliminary interpretations according to the various acoustic grammars. For example, the English language may be mapped into a detailed acoustic grammar representing the phonotactic rules of English, where words may be divided into syllables, which may further be divided into core components of an onset, a nucleus, and a coda, which may be further broken down into one or more sub-categories.

[025] Once the phonotactic rules of a speech signal have been identified, a detailed acoustic grammar tree may be constructed that accounts for the nuances of the speech signal. The acoustic grammar may include a loop of phonemes, or the phoneme loop may include a linking element to reduce a size of a search space associated with the grammar. Using the English language as an example, the grammar tree may include various branches representing English language

Google Exhibit 1019
Page 00577

syllables. The speech engine may traverse one or more grammar trees to generate one or more preliminary interpretations of a phoneme stream as a series of syllables that map to a word or phrase. By using phonemic recognition rather than word recognition, the size of the grammar can be reduced, which reduces the amount of time required to compile, load, and execute speech interpretation. Moreover, because the grammar maintains a high level of phonotactic constraints and therefore a large number of syllables, speech engine 112 may be very precise in generating phonemic representations of human verbalizations.

[026] An acoustic grammar used by speech engine 112 may be further optimized to reduce compile time, load time, and execution time by reducing the size of a search space associated with the acoustic grammar. Referring now to Fig. 2, a traditional grammar tree 120 is compared to an exemplary grammar tree according to one aspect of the invention to demonstrate the performance enhancements of speech engine 112. In traditional speech processing engines, nodes in a grammar tree 210 tend to represent words, or large-list applications may be supported provided through a grammar tree 210 where the nodes represent items in the large-list. This requires the speech engine to parse through a large number of transition states to come up with a recognition result, which degrades response time. An example of this is seen in the following grammar structure:

[027] "<street name> <city name>" → e.g., "NE 24th Street Bellevue"

[028] In the above example, a large list of street names is followed by a large list of city names. Assuming three elements in the list of street names, and three elements in the list of city names, this results in twenty-one transitions, which may be represented by traditional grammar tree 210. Every end-node of the first list is followed by all entries in the second list, potentially leading to very large grammars because most real-world large-list applications are likely to include much more than three list items. For example, a city may have hundreds or thousands of street names, and there may be hundreds or thousands of city names. Moreover, every element in the second segment of traditional grammar tree 210 is repeated, once for each first segment, which introduces redundancy.

[029] According to an aspect of the invention, the problems with traditional grammar trees may be resolved by using phonemic acoustic grammars instead of large-lists. The grammar may further be improved by including linking elements to reduce the number of transition states in the grammar. Thus, a grammar tree with a linking

Google Exhibit 1019
Page 00578

element 220 will merge after a first segment and then spread out again at a second segment, where the segments may represent a phoneme in an acoustic grammar, as discussed above. For example, assume a two-syllable word in an acoustic grammar consisting of three phonemes, which is able to reduce the number of transitions from twenty-one in a traditional grammar tree 210 to twelve in a grammar tree with a linking element 220. Two syllables and three phonemes are chosen to show the reduction in search space in a grammar tree with a linking element 220 as opposed to a corresponding traditional grammar tree 210, although a real-world acoustic grammar modeled after a language is likely to have a maximum of roughly fifty phonemes. Moreover, the search space may be further reduced by restricting available transitions based on phonotactic constraints for an acoustic model.

[030] Using the approach described in Fig. 2, adding a linking element to an acoustic grammar may reduce both grammar size and response time. Part of a speech signal may be mapped to the linking element in order to maintain the phonotactic rules of the acoustic grammar. The linking element may be an acoustic element that is likely to be triggered even if unpronounced. For example, a *schwa* represents an unstressed, central vowel in the English language (e.g., the first and last sound in the word "arena" is *schwa*). The phoneme *schwa* is an ideal linking element because of how it is represented in a frequency spectrum. That is, *schwa* is a brief sound and when a person opens their mouth to speak, there is a strong likelihood of passing through the frequencies of *schwa* even if unintended. Those skilled in the art will recognize that this approach may be extended to acoustic models of speech signals for other languages by using frequently elided phonemes as linking elements to reduce the search space of an acoustic grammar.

[031] Referring again to Fig. 1, speech engine 112 may receive a user verbalization and process the verbalization into a plurality of preliminary interpretations using the techniques described above. That is, the verbalization may be interpreted as a series of phonemes, and the series of phonemes may be mapped to one or more preliminary interpretations by traversing one or more acoustic grammars that are modeled after grammar 220 of Fig. 2. The plurality of preliminary interpretations may take the form of words, parts of words, phrases, utterances, or a combination thereof, and the plurality of preliminary interpretations may be arranged as a matrix, an array, or in another form. The plurality of preliminary interpretations are then

Google Exhibit 1019
Page 00579

passed to a speech sharpening engine 110 for deducing a most probable interpretation of the verbalization.

[032] According to various aspects of the invention, speech sharpening engine 110 may include an interpretation sharpening module 116, a policy module 114, an interpretation history analysis module 118, and a policy agent handler 120. The plurality of preliminary interpretations may be received by interpretation sharpening module 116, which forwards the preliminary interpretations to policy module 114 for further processing. Policy module 114 may include one or more context agents 126, one or more profile agents 128, and a context tracking module 130 that collectively revise the plurality of preliminary interpretations into a set of hypotheses that represent candidate recognitions of the verbalization. Policy module 114 may assign each hypothesis an interpretation score, and interpretation sharpening module 116 may designate the hypothesis with the highest (or lowest) interpretation score as a probable interpretation.

[033] According to one aspect of the invention, policy module 114 may include one or more context agents 126. Context agents 126 may represent domains of knowledge corresponding to a given context, such as song titles, city and street names, finance, movies, or other contexts. Context agents 126 may use context objects and associated dynamic languages to represent a corresponding context. Policy module 114 may also include one or more profile agents 128. Profile agents 128 may represent domains of knowledge corresponding to a given profile, such as a specific user, language, accent, or other profiles. Profile agents 128 may use profile objects and dynamic languages to represent a corresponding profile. Dynamic languages for context agents 126 or profile agents 128 may specify vocabularies, word combinations, phrases, sentence structures, criteria, and priority weightings for any given context or profile, respectively. The priority weightings may weight individual parameters according to one or more weighting factors, such as assigning a weight according to a frequency of use, a difficulty to understand, or other factors. Policy module 114 may also include a context-tracking module 130. Context tracking module 130 may track a verbalization context of a consecutive series of verbalizations. Context tracking module 130 may utilize one or more conversation trees to track the series of verbalizations. Context tracking sub-module 214 may track one or more past or current verbalization contexts of the series of verbalizations, and/or may make predictions regarding one or more future

Google Exhibit 1019
Page 00580

verbalization contexts of the series of verbalizations. Policy module 114 may utilize information about the verbalization context, generated by context tracking module 130, to generate one or more sharpened interpretations and corresponding interpretation scores.

[034] In some implementations, policy module 114 may use context tracking module 130 to apply objects from one or more context agents 126 and/or profile agents 128 to the preliminary interpretations provided by speech engine 112. The various agents may compete with each other using a set of heuristics in a phonetic fuzzy matcher, where an intent or context of the user may be identified based on the set of heuristics about how a request may be phrased in a given domain. A closest phonetic match may be identified for suspect words and/or phrases among the plurality of preliminary interpretations.

[035] The phonetic fuzzy matcher may include an M-Tree that is populated with context objects, profile objects, and/or dynamic language data from one or more of context agents 126 and/or profile agents 128. M-Trees are known to those skilled in the art. The M-Tree may assign relative priority weights to the context objects, profile objects, and/or dynamic language data in order to account for the possibility of misrecognized phonemes, extraneous phonemes, or erroneously deleted phonemes. A closest distance metric associated with the M-Tree may be used given the relative weightings of phoneme misrecognition, phoneme addition, and phoneme deletion for various contexts and/or profiles.

[036] According to one aspect of the invention, one or more passes may be taken over the plurality of preliminary interpretations to identify dominant words and/or phrases among the plurality of preliminary interpretations. Using the M-Tree weighted model, one or more candidate interpretations may be made based on relevant substitution of suspect words and/or phrases based on phonetic similarities and/or domain appropriateness. For example, if a set of dominant words appear to be a movie name, a candidate interpretation will substitute the relevant words and/or phrases to generate a candidate interpretation about movies. After a set of candidate interpretations have been generated, the candidate interpretations are analyzed using the M-Tree weighted model. With the relevant domains constrained by the candidate interpretations, a confidence or interpretation score may be assigned to each candidate interpretation, with the interpretation score representing a likelihood that a particular candidate interpretation is a correct interpretation of the

Google Exhibit 1019
Page 00581

verbalization. The candidate interpretations may then be returned to interpretation sharpening module 116, and interpretation sharpening module 116 may select a candidate interpretation with a highest (or lowest) interpretation score as a probable interpretation of the verbalization.

[037] According to various implementations of the invention, speech sharpening engine 110 may include an interpretation history analysis module 118. Interpretation history analysis module 118 may include an information storage module 122 a frequency module 124. Information storage module 122 may store information related to verbalizations, including components of verbalizations, preliminary interpretations, dominant words and/or phrases, candidate interpretations, probable interpretations, and/or interpretation scores associated with verbalizations, as well as whether or not a verbalization was interpreted correctly, or other information. Interpretation history analysis module 118 may also include a frequency module 124. Frequency module 124 may use some or all of the information stored in information storage module 122 to generate one or more frequencies related to one or more past verbalizations. For example, frequency module 124 may calculate a word usage frequency, a word combination frequency, a frequency related to a set of verbalizations that are phonetically similar but have distinct meanings, an interpretation error frequency for a particular verbalization, or other frequencies.

[038] Information stored and/or generated by interpretation history analysis module 118 may be used to augment speech sharpening engine 110. In some implementations, the information may be used to adjust various weights used in phonetic models, such as context agents 126 or profile agents 128, as well as adapting the relative weights in the M-Tree in context tracking module 130 to enhance accuracy for subsequent verbalizations. In another implementation, the stored information may be sent to a third-party or commercial entity for analyzing the data and developing new domain agents or further improving the accuracy of speech sharpening engine 110. For example, one or more parameters and/or weighting schemes of an agent may be augmented based on a frequency generated by interpretation history analysis module 118. Other augmentations related to information stored on and/or generated by interpretation history analysis module 118 may be made. Speech sharpening engine 110 may also include a policy agent handler 120 that may augment, update, remove, and/or provide one or more domain agents to policy module 114. A domain agent may include one or more new,

Google Exhibit 1019
Page 00582

modified, or updated context agents 126 and/or profile agents 128. Policy agent handler 120 may also augment or update the M-Tree in context tracking module 130 to adjustments in priority weighting schemes or phonetic models. Policy agent handler 120 may be controlled, directly or indirectly, by a third party, such as a commercial entity, and domain agents may be augmented, updated, removed, and/or provided by policy agent handler 120 as part of a commercial agreement, licensing agreement, subscription agreement, maintenance agreement, or other agreement.

[039] Referring to Fig. 3, a flow chart demonstrating an exemplary method for enhancing the performance and accuracy of speech interpretation is provided. The method may begin by receiving a user verbalization at an operation 312. The received user verbalization may be electronically captured at operation 312, such as by a microphone or other electronic audio capture device. The electronically captured verbalization may be provided to a speech interpretation engine, such as speech engine 112 in Fig. 1.

[040] The speech interpretation may then generate one or more preliminary interpretations of the received verbalization at an operation 314. According to one implementation of the invention, the plurality of preliminary interpretations may be generated using phonetic dictation, grammar trees with linking elements, or any combination thereof to improve performance and enhance accuracy. Phonetic dictation and reducing a search space of a grammar tree by including linking elements is discussed in greater detail above. The preliminary interpretations may be arranged in any predetermined form, such as an array, a matrix, or other forms.

[041] In an operation 320, the preliminary interpretations may be provided to a speech sharpening engine. The speech sharpening engine may take one or more passes over the plurality of preliminary interpretations to identify dominant words and/or phrases in operation 320. This information may then be used to generate one or more candidate interpretations. The candidate interpretations may be based on various domain agents, such as context agents and/or profile agents, which may be organized as a weighted domain model, such as an M-Tree. For example, if a set of dominant words sound like a movie name, apply policies operation 320 may generate a candidate interpretation that substitutes relevant words and/or phrases based on a domain agent populated with movie titles. Additional passes may be made over the candidate interpretations, which may be constrained by domain

Google Exhibit 1019
Page 00583

information associated with the candidate interpretations, to thereby generate a confidence score or interpretation score for each candidate interpretation. The interpretation score may represent a likelihood that a particular candidate interpretation is a correct interpretation of the verbalization received in operation 312. The operation of apply policies 320 is described in greater detail above in reference to Fig. 1.

[042] The candidate interpretations and corresponding interpretation scores may then be analyzed to determine a probable interpretation in an operation 322. In one implementation of the invention, a candidate interpretation with a highest (or lowest) score may be designated as a probable interpretation. The probable interpretation may then be output in an operation 324, such as for use in a voice-activated vehicular navigation system, a voice-controlled server or desktop computer, or other electronic device that can be controlled using voice commands.

[043] Information relating to the verbalization and the interpretations of the verbalization may be provided in a store interpretation operation 325. Store interpretation operation 324 may store interpretation information related to verbalizations, such as components of verbalizations, preliminary interpretations, dominant words and/or phrases, candidate interpretations, probable interpretations, and/or interpretation scores associated with verbalizations, as well as whether or not a verbalization was interpreted correctly, or other information. In some implementations of the invention, some or all of the interpretation information stored at store interpretation operation 324 may be used to determine one or more frequencies at a determine frequencies operation 326. The frequencies calculated at determine frequencies operation 326 may include one or more frequencies related to past verbalizations, such as, a word usage frequency, a word combination frequency, a frequency related to a set of verbalizations that are phonetically similar but have distinct meanings, an interpretation error frequency for a particular verbalization, or other frequencies. Determine frequencies operation 326 may be performed by interpretation history analysis module 118.

[044] In various implementations, a decision may be made whether to augment a speech sharpening engine in an augmentation decision operation 328. The decision concerning system augmentation may be based at least in part on information generated at determine frequencies block 326, such as one or more frequencies, or other information. If it is decided that no augmentation is needed, no further action is

Google Exhibit 1019
Page 00584

taken until another verbalization is captured, and the method ends. In some instances, decision operation 328 may determine that augmentation may be made and control passes to an augment system operation 330. Augment system operation 330 may include making an augmentation to a speech sharpening engine. For example, one or more domain agents may be augmented to reflect probabilities of an interpretation being a correct interpretation of a verbalization, to update a user profile, or other augmentation. Dynamic languages associated with context agents and/or profile agents may be augmented, or parameters weights may be augmented to enhance accuracy when interpreting subsequent verbalizations. For example, an adaptive misrecognition technique may adjust the various weights in a phonetic model or update similarity weights for regional accents, or other augmentations may be made. In parallel to augment system operation 330, new agent policies may be received in an operation 332. For example, a third party or commercial entity may redesign or modify various domain agents, new domain agents may be developed and installed as plug-ins, domain agents that are unreliable may be removed, or other augmentations or modifications may be made. Thus, the method continually refines the domain agents and the weighting of various parameters in order to refine the accuracy of the speech sharpening engine for subsequent verbalizations.

[045] The above disclosure has been described in terms of specific exemplary aspects, implementations, and embodiments of the invention. However, those skilled in the art will recognize various changes and modifications that may be made without departing from the scope and spirit of the invention. For example, references throughout the specification to "one implementation," "one aspect," "an implementation," or "an aspect" may indicate that a particular feature, structure, or characteristic is included in at least one implementation. However, the particular features, structures, or characteristics may be combined in any suitable manner in one or more implementations. Therefore, the specification and drawings are to be regarded as exemplary only, and the scope of the invention is to be determined solely by the appended claims.

Google Exhibit 1019
Page 00585

CLAIMS

What is claimed is:

1. A computer-implemented method for interpreting speech, the method comprising:

receiving a user verbalization;

parsing the verbalization into a plurality of phonemes; and

generating at least one interpretation of the verbalization based on the plurality of phonemes.

2. The method of claim 1, wherein generating the at least one preliminary interpretation includes traversing an acoustic grammar based on the plurality of phonemes.

3. The method of claim 2, wherein the acoustic grammar is a phonemic representation of a speech signal.

4. The method of claim 3, wherein the acoustic grammar includes a linking element between segments, the linking element being an acoustic element in the speech signal.

5. The method of claim 4, wherein the acoustic element is a phoneme representing an unstressed vowel.

6. The method of claim 1, further comprising:

revising the at least one interpretation, wherein the revising includes generating a plurality of candidate interpretations of the verbalization;

assigning an interpretation score to each candidate interpretation; and

selecting a candidate interpretation with a highest interpretation score as a probable interpretation of the verbalization.

7. The method of claim 6, wherein revising the at least one interpretation includes:

applying the at least one interpretation against at least one domain agent;

identifying suspect words and/or phrases in the at least one interpretation;

identifying a closest phonetic match for the suspect words and/or phrases based on a relative weighting of parameters in at least one domain associated with the at least one domain agent; and

substituting the closest phonetic match for the suspect words and/or phrases to generate a candidate interpretation.

Google Exhibit 1019
Page 00586

8. The method of claim 7, wherein the closest phonetic match is stored in an M-Tree, and wherein identifying the closest phonetic match includes using a closest-distance metric associated with the M-Tree.

9. The method of claim 7, wherein the parameters include at least one of a context of the verbalization, a verbalization criteria, a history of user verbalizations, a user profile, or domain specific information.

10. The method of claim 7, further comprising adjusting the relative weighting of parameters in response to the selection of a probable interpretation of the verbalization.

11. A method for sharpening a preliminary interpretation of a user verbalization, the preliminary interpretation generated by a speech interpretation engine, the method comprising:

revising the preliminary interpretation, wherein the revising includes generating a plurality of candidate interpretations of the verbalization;

assigning an interpretation score to each candidate interpretation; and

selecting a candidate interpretation with a highest interpretation score as a probable interpretation of the verbalization.

12. The method of claim 11, wherein revising the preliminary interpretation includes:

applying the preliminary interpretation against at least one domain agent;

identifying suspect words and/or phrases in the preliminary interpretation;

identifying a closest phonetic match for the suspect words and/or phrases based on a relative weighting of parameters in at least one domain associated with the at least one domain agent; and

substituting the closest phonetic match for the suspect words and/or phrases to generate a candidate interpretation.

13. The method of claim 12, wherein the closest phonetic match is stored in an M-Tree, and wherein identifying the closest phonetic match includes using a closest-distance metric associated with the M-Tree.

14. The method of claim 12, wherein the parameters include at least one of a context of the verbalization, a verbalization criteria, a history of user verbalizations, a user profile, or domain specific information.

Google Exhibit 1019
Page 00587

15. The method of claim 12, further comprising adjusting the relative weighting of parameters in response to the selection of a probable interpretation of the verbalization.

16. A system for interpreting speech, the system comprising:

a device that receives a user verbalization and generates an electronic signal corresponding to the user verbalization; and

a speech interpretation engine that receives the electronic signal corresponding to the user verbalization, wherein the speech interpretation engine parses the signal into a plurality of phonemes and generates at least one interpretation of the verbalization based on the plurality of phonemes.

17. The system of claim 16, wherein generating the at least one preliminary interpretation includes traversing an acoustic grammar based on the plurality of phonemes.

18. The system of claim 17, wherein the acoustic grammar is a phonemic representation of a speech signal.

19. The system of claim 18, wherein the acoustic grammar includes a linking element between segments, the linking element being an acoustic element in the speech signal.

20. The system of claim 19, wherein the acoustic element is a phoneme representing an unstressed vowel.

21. The system of claim 16, further comprising a speech sharpening engine configured to:

revise the at least one interpretation by generating a plurality of candidate interpretations of the verbalization;

assign an interpretation score to each candidate interpretation; and

select a candidate interpretation with a highest interpretation score as a probable interpretation of the verbalization.

22. The system of claim 21, wherein the speech sharpening engine is configured to revise the at least one interpretation by:

applying the at least one interpretation against at least one domain agent;

identifying suspect words and/or phrases in the at least one interpretation;

identifying a closest phonetic match for the suspect words and/or phrases based on a relative weighting of parameters in at least one domain associated with the at least one domain agent; and

Google Exhibit 1019
Page 00588

substituting the closest phonetic match for the suspect words and/or phrases to generate a candidate interpretation.

23. The system of claim 22, wherein the closest phonetic match is stored in an M-Tree, and wherein identifying the closest phonetic match includes using a closest-distance metric associated with the M-Tree.

24. The system of claim 22, wherein the parameters include at least one of a context of the verbalization, a verbalization criteria, a history of user verbalizations, a user profile, or domain specific information.

25. The system of claim 22, wherein the speech sharpening engine is further configured to adjust the relative weighting of parameters in response to the selection of a probable interpretation of the verbalization.

26. A system for sharpening a preliminary interpretation of a user verbalization received by an electronic capture device, the preliminary interpretation generated by a speech interpretation engine, the system comprising a speech sharpening engine containing at least one processor configured to:

revise the preliminary interpretation by generating a plurality of candidate interpretations of the verbalization;

assign an interpretation score to each candidate interpretation; and

select a candidate interpretation with a highest interpretation score as a probable interpretation of the verbalization.

27. The system of claim 26, wherein the speech sharpening engine is further configured to revise the preliminary interpretation by:

applying the preliminary interpretation against at least one domain agent;

identifying suspect words and/or phrases in the preliminary interpretation;

identifying a closest phonetic match for the suspect words and/or phrases based on a relative weighting of parameters in at least one domain associated with the at least one domain agent; and

substituting the closest phonetic match for the suspect words and/or phrases to generate a candidate interpretation.

28. The system of claim 27, wherein the closest phonetic match is stored in an M-Tree, and wherein identifying the closest phonetic match includes using a closest-distance metric associated with the M-Tree.

Google Exhibit 1019
Page 00589

29. The system of claim 27, wherein the parameters include at least one of a context of the verbalization, a verbalization criteria, a history of user verbalizations, a user profile, or domain specific information.

30. The system of claim 27, wherein the speech sharpening engine is further configured to adjust the relative weighting of parameters in response to the selection of a probable interpretation of the verbalization.

31. A computer-readable medium containing computer-executable instructions for interpreting speech, the computer-executable instructions configured to:

receive a user verbalization;

parse the verbalization into a plurality of phonemes; and

generate at least one interpretation of the verbalization based on the plurality of phonemes.

32. The computer-readable medium of claim 31, wherein the computer-executable instructions are configured to generate the at least one preliminary interpretation by traversing an acoustic grammar based on the plurality of phonemes.

33. The computer-readable medium of claim 32, wherein the acoustic grammar is a phonemic representation of a speech signal.

34. The computer-readable medium of claim 33, wherein the acoustic grammar includes a linking element between segments, the linking element being an acoustic element in the speech signal.

35. The computer-readable medium of claim 34, wherein the acoustic element is a phoneme representing an unstressed vowel.

36. The computer-readable medium of claim 31, wherein the computer-executable instructions are further configured to:

revise the at least one interpretation by generating a plurality of candidate interpretations of the verbalization;

assign an interpretation score to each candidate interpretation; and

select a candidate interpretation with a highest interpretation score as a probable interpretation of the verbalization.

37. The computer-readable medium of claim 36, wherein the computer-executable instructions are configured to revise the at least one interpretation by:

applying the at least one interpretation against at least one domain agent;

identifying suspect words and/or phrases in the at least one interpretation;

Google Exhibit 1019
Page 00590

identifying a closest phonetic match for the suspect words and/or phrases based on a relative weighting of parameters in at least one domain associated with the at least one domain agent; and

substituting the closest phonetic match for the suspect words and/or phrases to generate a candidate interpretation.

38. The computer-readable medium of claim 37, wherein the closest phonetic match is stored in an M-Tree, and wherein identifying the closest phonetic match includes using a closest-distance metric associated with the M-Tree.

39. The computer-readable medium of claim 37, wherein the parameters include at least one of a context of the verbalization, a verbalization criteria, a history of user verbalizations, a user profile, or domain specific information.

40. The computer-readable medium of claim 37, wherein the computer-executable instructions are further configured to adjust the relative weighting of parameters in response to the selection of a probable interpretation of the verbalization.

41. A computer-readable medium containing computer-executable instructions for sharpening a preliminary interpretation of a user verbalization, the preliminary interpretation generated by a speech interpretation engine, the computer-executable instructions configured to:

revise the preliminary interpretation by generating a plurality of candidate interpretations of the verbalization;

assign an interpretation score to each candidate interpretation; and

select a candidate interpretation with a highest interpretation score as a probable interpretation of the verbalization.

42. The computer-readable medium of claim 41, wherein the computer-executable instructions are configured to revise the preliminary interpretation by:

applying the preliminary interpretation against at least one domain agent;

identifying suspect words and/or phrases in the preliminary interpretation;

identifying a closest phonetic match for the suspect words and/or phrases based on a relative weighting of parameters in at least one domain associated with the at least one domain agent; and

substituting the closest phonetic match for the suspect words and/or phrases to generate a candidate interpretation.

Google Exhibit 1019
Page 00591

43. The computer-readable medium of claim 42, wherein the closest phonetic match is stored in an M-Tree, and wherein identifying the closest phonetic match includes using a closest-distance metric associated with the M-Tree.

44. The computer-readable medium of claim 42, wherein the parameters include at least one of a context of the verbalization, a verbalization criteria, a history of user verbalizations, a user profile, or domain specific information.

45. The computer-readable medium of claim 42, wherein the computer-executable instructions are further configured to adjust the relative weighting of parameters in response to the selection of a probable interpretation of the verbalization.

Google Exhibit 1019
Page 00592

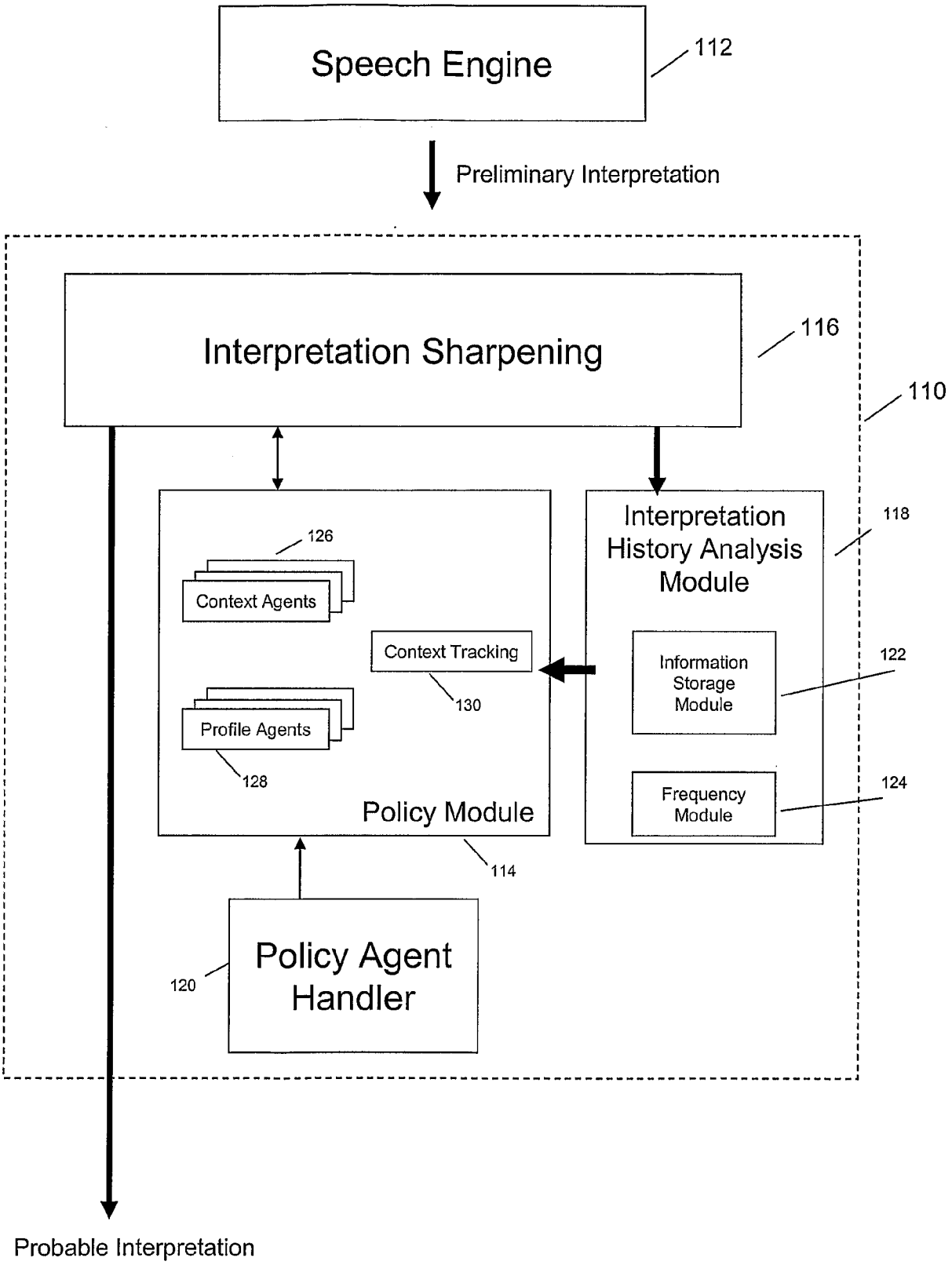
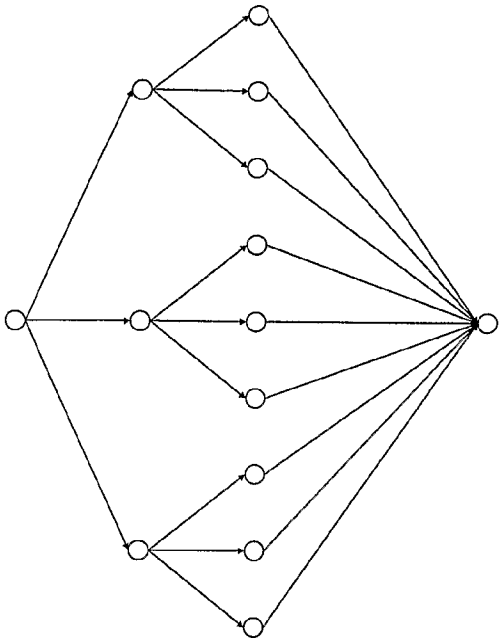

Speech Engine
112
Preliminary Interpretation
Interpretation Sharpening
116
110
Interpretation History Analysis Module
118
126
Context Agents
Context Tracking
130
Information Storage Module
122
Profile Agents
128
Frequency Module
124
Policy Module
114
Policy Agent Handler
120
Probable Interpretation

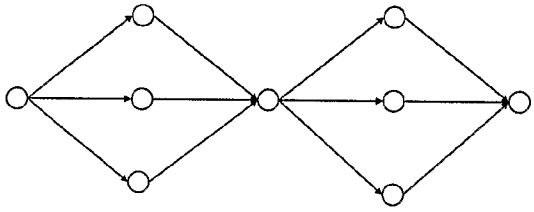
Fig. 1

Traditional Grammar Tree



210

Grammar Tree with Linking Element



220

Figure 2

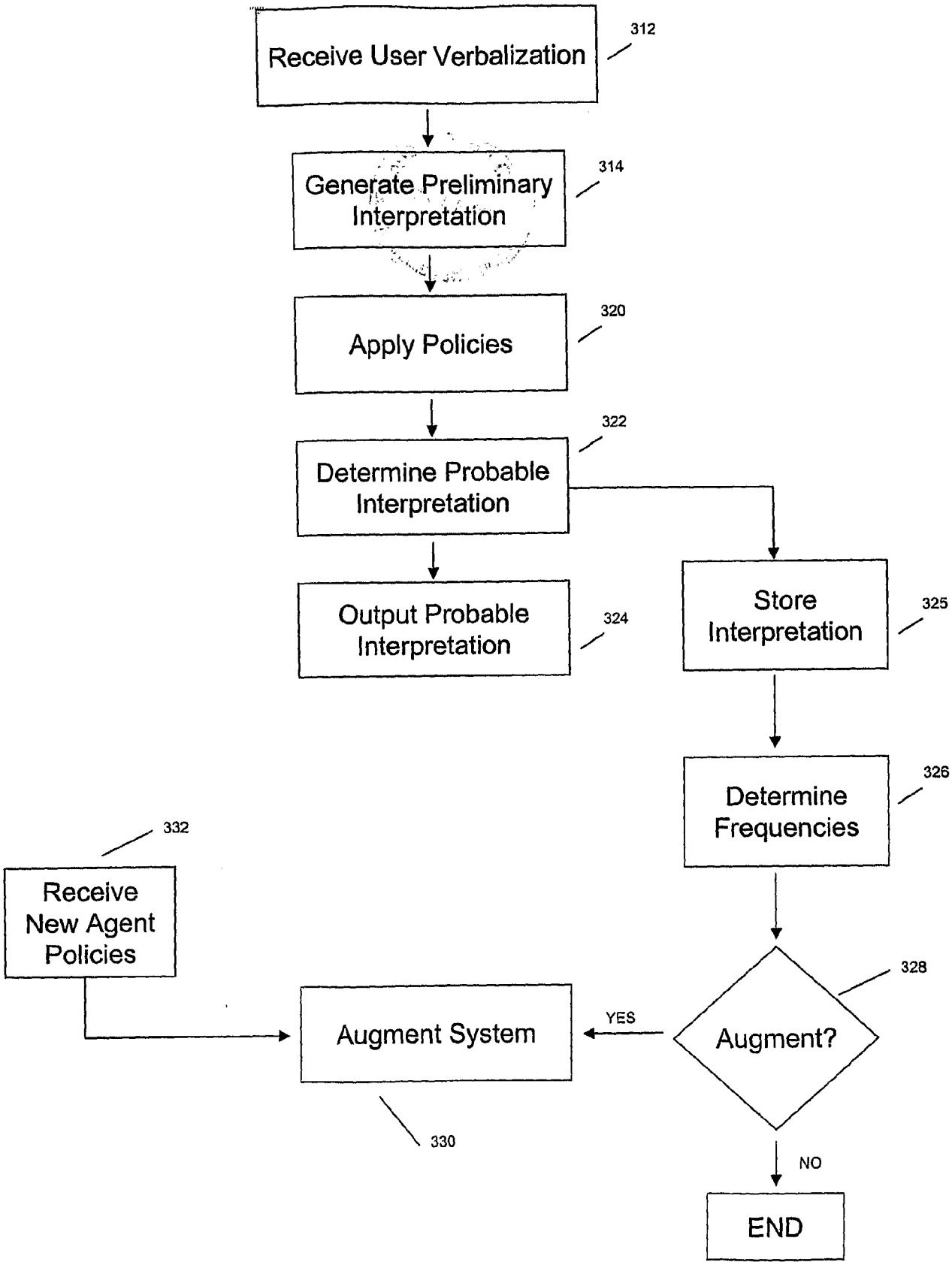

Receive User Verbalization
312
Generate Preliminary Interpretation
314
Apply Policies
320
Determine Probable Interpretation
322
Output Probable Interpretation
324
Store Interpretation
325
Determine Frequencies
326
Augment?
328
YES
Augment System
330
NO
END
Receive New Agent Policies
332

Fig. 3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2008 (14.08.2008)

PCT

(10) International Publication Number
WO 2008/098039 A1

(51) International Patent Classification:
G06Q 30/00 (2006.01)

(21) International Application Number:
PCT/US2008/053157

(22) International Filing Date: 6 February 2008 (06.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/671,526 6 February 2007 (06.02.2007) US

(71) Applicant *(for all designated States except US)*: **VOICE-BOX TECHNOLOGIES, INC.** [US/US]; 11980 NE 24th Street, Bellevue, WA 98005 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FREEMAN, Tom** [US/US]; 6735 83rd Avenue, S.E., Mercer Island, WA 98040 (US). **KENNEWICK, Mike** [US/US]; 13407 Ne 37th Place, Bellevue, WA 98005 (US).

(74) Agents: **GATTO, James, G.** et al.; Pillsbury Winthrop Shaw Pittman LLP, P.O. Box 10500, McLean, VA 22102 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(54) **Title:** SYSTEM AND METHOD FOR SELECTING AND PRESENTING ADVERTISEMENTS BASED ON NATURAL LANGUAGE PROCESSING OF VOICE-BASED INPUT

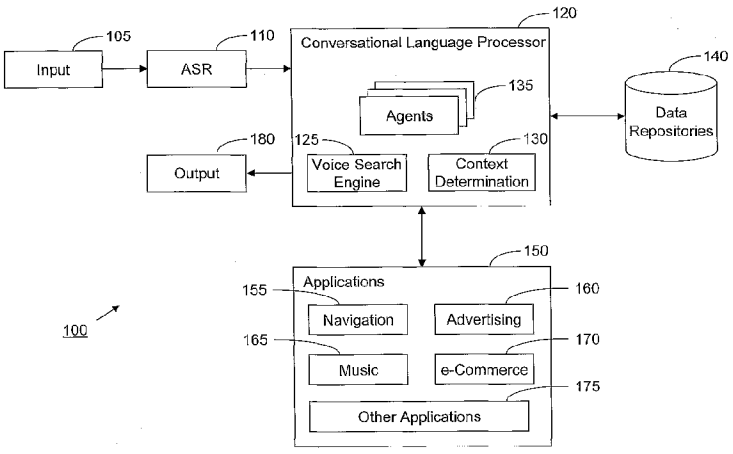


Fig. 1

(57) **Abstract:** A system and method for selecting and presenting advertisements based on natural language processing of voice-based inputs is provided. A user utterance may be received at an input device, and a conversational, natural language processor may identify a request from the utterance. At least one advertisement may be selected and presented to the user based on the identified request. The advertisement may be presented as a natural language response, thereby creating a conversational feel to the presentation of advertisements. The request and the user's subsequent interaction with the advertisement may be tracked to build user statistical profiles, thus enhancing subsequent selection and presentation of advertisements.

WO 2008/098039 A1

SYSTEM AND METHOD FOR SELECTING AND PRESENTING ADVERTISEMENTS BASED ON NATURAL LANGUAGE PROCESSING OF VOICE-BASED INPUT

FIELD OF THE INVENTION

[001] The present invention relates to selecting and presenting advertisements based on natural language processing of voice-based input.

BACKGROUND OF THE INVENTION

[002] As technology advances, consumer electronics devices tend to play larger roles due to increased functionality and mobility. For example, mobile phones, navigation devices, embedded devices, and other such devices provide a wealth of functionality beyond core applications. However, increased functionality adds difficulty to the learning curve associated with using electronic devices, and increased mobility intensifies the demand for simple mechanisms to interact with devices on the go. For example, existing systems tend to have complex human to machine interfaces, which may inhibit mass-market adoption for various technologies. For example, when a user wishes to perform a relatively simple task on a mobile phone, such as purchasing a ring tone, the user often is forced to navigate through a series of menus and press a series of buttons. In some instances, this may result in the transaction not necessarily occurring, as the user may prefer to avoid the hassles altogether. As such, there is ever-growing demand for ways to exploit technology in intuitive ways.

[003] Voice recognition software may enable a user to exploit applications and features of a device that may otherwise be unfamiliar, unknown, or difficult to use. However, many existing voice user interfaces (when they actually work) still require significant learning on the part of the user. For example, users often cannot directly issue a request for a system to retrieve information or perform an action without having to memorize specific syntaxes, words, phrases, concepts, semantic indicators, or other keywords/qualifiers. Similarly, when users are uncertain of particular needs, many existing systems do not engage the user in a productive, cooperative dialogue to resolve requests and advance a conversation. Instead, many existing speech interfaces force users to use a fixed set commands or

keywords to communicate requests in ways that systems can understand. Using existing voice user interfaces, there is virtually no option for dialogue between the user and the system to satisfy mutual goals.

[004] The lack of adequate voice user interfaces results in missed opportunities for providing valuable and relevant information to users. Not only does this potentially leave user requests unresolved, in certain instances, providers of goods and services may lose out on potential business. In an increasingly global marketplace, where marketers are continually looking for new and effective ways to reach consumers, the problems with existing voice user interfaces leaves a large segment of consumer demand unfulfilled. Furthermore, existing techniques for marketing, advertising, or otherwise calling consumers to action fail to effectively utilize voice-based information, which is one of the most natural, intuitive methods of human interaction.

[005] Existing systems suffer from these and other problems.

SUMMARY OF THE INVENTION

[006] According to various aspects of the invention, a system and method for selecting and presenting advertisements based on natural language processing of voice-based inputs is provided. A natural language voice-based input may be received by a voice user interface. The voice-based input may include a user utterance, and a request may be identified from the utterance. Appropriate action may be taken to service the request, while one or more advertisements may be selected and presented to the user. Advertisements may be selected based on various criteria, including content of the input (e.g., concepts, semantic indicators, etc.), an activity related to the input (e.g., a relation to a request, a requested application, etc.), user profiles (e.g., demographics, preferences, location, etc.), or in other ways. A user may subsequently interact with the advertisement (e.g., via a voice-based input), and action may be taken in response to the interaction. Furthermore, the interaction may be tracked to build statistical profiles of user behavior based on affinities or clusters among advertisements, user profiles, contexts, topics, semantic indicators, concepts, or other criteria.

[007] According to various aspects of the invention, advertisers may create advertisements, which may be stored in an advertisement repository. For example,

Google Exhibit 1019
Page 00598

advertisements may include sponsored messages, calls to action, purchase opportunities, trial downloads, or any other marketing communication, as would be apparent to those skilled in the art. Advertisers may specify various parameters to associate with the advertisements, such as various contexts or topic concepts (e.g., semantic indicators for a "music" concept may include words such as "music," "tunes," "songs," etc.), target demographics (e.g., a preferred audience), marketing criteria or prices for insertion (e.g., dynamic or static pricing based on various marketing criteria), or other information, as would be apparent. The advertisement repository may be associated with a server, where in response to a voice-based input from a user (e.g., at a voice-enabled device), a communications link may be established with the server. Information may be extracted from the voice-based input (e.g., words in the input, applications requested by the input, etc.), and the extracted information may be correlated with user profiles, advertisement parameters, or other information to determine which advertisements to select in relation to the voice-based input. The server may subsequently communicate the selected advertisements to the user, and the server may track the user's subsequent interaction with the selected advertisements.

[008] Other objects and advantages of the invention will be apparent based on the following drawings and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[009] Fig. 1 illustrates a block diagram of an exemplary system for implementing a voice user interface according to various aspects of the invention.

[010] Fig. 2 illustrates a block diagram of an exemplary advertising system according to various aspects of the invention.

[011] Fig. 3 illustrates a flow diagram of an exemplary method for selecting and presenting advertisements based on voice-based inputs according to various aspects of the invention.

DETAILED DESCRIPTION

[012] Referring to Fig. 1, an exemplary system 100 for implementing a voice user interface is illustrated according to various aspects of the invention. System 100 may enable users to perform various tasks on a voice-enabled device. For example, users may control navigation

Google Exhibit 1019
Page 00599

devices, media devices, personal computers, personal digital assistants, or any other device supporting voice-based inputs. System 100 may enable users to request voice-enabled devices to retrieve information or perform various tasks, among other things, using natural language voice-based inputs. For example, system 100 may interpret natural language voice-based inputs and generate responses using, among other things, techniques described in co-pending U.S. Patent Application Ser. No. 10/452,147, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed June 3, 2003; and U.S. Patent Application Ser. No. 10,618,633, entitled "Mobile Systems and Methods for Responding to Natural Language Speech Utterance," filed June 15, 2003, both of which are hereby incorporated by reference in their entirety.

[013] According to various aspects of the invention, system 100 may receive a user input, including at least a voice-based user utterance, at an input device 105. Input device 105 may include any suitable device, or combination of devices, for receiving a voice-based input (e.g., a microphone). In various implementations, input device 105 may include a multi-modal input, such as a touch-screen interface, keypad, or other input. The received utterance may be processed by an Automatic Speech Recognizer 110. Automatic Speech Recognizer 110 may generate one or more preliminary interpretations of the utterance using various techniques. For example, Automatic Speech Recognizer 110 may interpret the utterance using techniques of phonetic dictation to recognize a stream of phonemes. Further, Automatic Speech Recognizer 110 may perform post-processing to enhance the preliminary interpretations. For example, Automatic Speech Recognizer 110 may vary interpretations of an utterance, or components of an utterance, from one context to another. Other techniques for enhancing an interpretation of a user utterance may be used, such as those described in co-pending U.S. Patent Application Ser. No. 11/513,269, entitled "Dynamic Speech Sharpening," filed August 31, which is hereby incorporated by reference in its entirety.

[014] The one or more preliminary interpretations may be provided to a conversational language processor 120. Conversational language processor 120 may include a voice search engine 125, a context determination module 130, and one or more agents 135, among other things, to enable cooperative, conversational interaction between the user and system 100. Conversational language processor 120 may be communicatively coupled to one or more data repositories 140 and one or more applications 150. Conversational language processor

Google Exhibit 1019
Page 00600

120 may generate a domain-specific conversational response, which may be returned to the user as an output 180. Output 180 may include a multi-modal output (e.g., by simultaneously returning a voice-based response and displaying information on a display device).

[015] System 100 may further include an interaction with one or more applications 150 to service one or more requests in the utterance. For example, the utterance may include one or more requests for performing an action, retrieving information, or various combinations thereof. Output 180 may include a conversational response to advance a conversation to service requests by invoking one or more applications 150, as appropriate. For example, applications 150 may include a navigation application 155, an advertising application 160, a music application, an electronic commerce application 170, and/or other applications 175. Furthermore, Automatic Speech Recognizer 110, conversational language processor 120, data repositories 140, and/or applications 150 may reside locally (e.g., on a user device), remotely (e.g., on a server), and/or hybrid local/remote processing models may be used (e.g., lightweight applications may be processed locally while computationally intensive applications may be processed remotely).

[016] Conversational language processor 120 may build long-term and/or short-term shared knowledge in one or more knowledge source. For example, shared knowledge sources may include information about previous utterances, requests, and other user interactions to inform generating an appropriate response to a current utterance. The shared knowledge may include public/non-private (i.e., environmental) knowledge, as well as personal/private (i.e., historical) knowledge. For example, conversational language processor 120 may use context determination module 130 to establish a context for a current utterance by having domain agents 135 competitively generate a context-based interpretation of the utterance (e.g., by scoring possible interpretations and selecting a highest scoring interpretation). As such, agents 135 may model various domains (e.g., navigation, music, a specific user, global users, advertising, e-commerce, etc.), and conversational language processor 120 may interpret and/or respond to a voice-based input accordingly. For example, context-based interpretations and responses to a voice-based input may be generated using techniques described in co-pending U.S. Patent Application Ser. No. 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed August 5, 2005; and U.S. Patent Application Ser. No. 11/212,693,

Google Exhibit 1019
Page 00601

entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," filed August 29, 2005, both of which are hereby incorporated by reference in their entirety.

[017] Furthermore, conversational language processor 120 may support adaptive misrecognition to reinterpret a current utterance and/or one or more previous utterances. For example, information contained in a current utterance may indicate that interpretations for one or more previous utterances were incorrect, and therefore, the previous utterances may be reinterpreted to improve subsequent interpretations. Accordingly, conversational language processor 120 may use the techniques described herein, along with various other techniques, to interpret and respond to conversational, natural language utterances. Conversational language processor 120 may use various other techniques as will be apparent, such as those described in co-pending U.S. Patent Application Ser. No. 11/200,164, entitled "System and Method of Supporting Adaptive Misrecognition in Conversational Speech," filed August 10, 2005; and U.S. Patent Application Ser. No. 11/580,926, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed October 16, 2006, both of which are hereby incorporated by reference in their entirety.

[018] Referring to Fig. 2, an exemplary advertising system 200 is illustrated according to various aspects of the invention. System 200 may include a server 230 for receiving one or more advertisements from an advertiser 220, wherein the advertisements may be stored in a data repository 260 associated with server 230. For example, advertisements may include sponsored messages or marketing communications, calls to action, purchase opportunities, trial downloads, coupons, or any other suitable marketing, advertising, campaign, or other information, as would be apparent to those skilled in the art. A voice-enabled device 210 may receive a voice-based input and establish communications with advertising server 230. Subsequently, advertising server 230 may select one or more advertisements from among the advertisements stored in data repository 260, and the selected advertisements may be provided to the voice-enabled device for presentation to a user.

[019] Advertiser 220 may access advertising server 230 via an advertiser interface 245. Advertisers 220 may upload targeted advertisements to server 230 via advertiser interface 245, and server 230 may store the advertisements in data repository 260. The advertisements may include graphically-based advertisements that include banners, images,

Google Exhibit 1019
Page 00602

audio, video, or any suitable combination thereof. Furthermore, the advertisements may include interactive or embedded information, such as links, metadata, or computer-executable instructions, or any suitable combination thereof. Advertisers may specify criteria for a campaign or targeting information for an advertisement (e.g., a start date, an end date, budget information, geo-targeting information, conceptual or contextual information, or any other suitable criteria), which may be used to facilitate selecting an advertisement in relation to a particular voice-based input.

[020] In addition to providing interface 245 for advertisers, server 230 may include a content/action identification module 235, a user profile module 240, an advertisement selection module 250, and a tracking module 255. Users may submit voice-based requests to voice-enabled device 210, and voice-enabled device 210 may communicate information about the voice-based input to server 230. Server 230 may invoke advertisement selection module 250 to extract relevant information from about the voice-based input, where advertisement selection module 250 may select one or more advertisements relevant to the voice-based input based on information extracted using content/action identification module 235 and/or user profile module 240.

[021] For example, content/action identification module 235 may identify content of the voice-based input (e.g., words in the input), requested information (e.g., search results, a web page, music, video, graphics, or other information), requested actions (e.g., calculating a navigation route, placing a telephone call, playing a song, etc.), a category or topic related to the input (e.g., music, business, stocks, sports, navigation, movies, etc.), or other criteria to use in selecting an advertisement. Further, user profile module 240 may identify characteristics of a specific user (e.g., demographics, personal preferences, location-based information, etc.), global user profiles (e.g., demographic profiles, click-through rates, etc.), or other criteria to use in selecting an advertisement. Moreover, advertisement selection module 250 may account for where a request originates from. For example, advertisements may be selected based on a default user location (e.g., identified from a user profile), current geolocation information (e.g., identified from a navigation device), whether an affiliate or partner of server 230 initiated the request, or other criteria.

[022] For instance, a user may request airline reservations via voice-enabled device 210, and content/action identification module 235 may identify specific words used in the request, a category related to the request (e.g., travel, airlines, hotels, etc.), or other

Google Exhibit 1019
Page 00603

information. Furthermore, user profile module 240 may identify relevant characteristics of the user (e.g., user-specific demographics, location information, preferred airlines or hotels, etc.), as well as global user characteristics (e.g., most popular airlines). In various implementations, advertisements may be selected by assigning a score to each advertisement (e.g., based on click-through rates, relevance metrics, target audiences, etc.). As such, advertisement selection module 250 may correlate the information about the request to select advertisements stored in data repository 260, and server 230 may communicate the selected advertisements to voice-enabled device 210. Furthermore, selected advertisements may be presented according to a predetermined ordering or ranking (e.g., based on a ranking of relevance to an advertisement).

[023] In various implementations, advertisement selection module 250 may retrieve a predetermined number of advertisements for any given request. Furthermore, the selected advertisements may depend upon a presentation format. For example, advertisements may be selected based on an amount of available space on a display of voice-enabled device 210 and/or a size/shape of the selected advertisements. In another example, voice-based advertisements may be selected and presented to the user audibly (e.g., a "hands-free" advertisement may be preferred when voice-enabled device 210 is a telematics device).

[024] Furthermore, the user's subsequent interaction with an advertisement may be tracked using tracking module 255. For example, tracking module 255 may determine whether a conversion or click-through occurs for each advertisement presented to users. Further, tracking module 255 may maintain accounting and/or billing information associated with advertisers 220. For example, advertisers 220 may specify a maximum insertion cost, a cost-per-click-through, an average insertion cost, or other criteria specifying a budget constraint for an advertisement. As such, tracking module 255 may track which advertisements are selected and/or presented, which advertisements result in a conversion or click-through, whether a click-through or conversion results in a transaction or sale, associations between advertisements and users, requests, concepts, semantic indicators, and/or other criteria. For example, tracking user interaction with advertisements may be used to build user-specific and/or global statistical profiles that map or cluster advertisements to topics, semantic indicators, contexts, concepts, etc. based on user behavior, demographics, targeting constraints, content of advertisements, content of requests, actions associated with requests, or other statistically relevant information.

Google Exhibit 1019
Page 00604

Accordingly, the tracking information may be used to bill or invoice advertisers 220, as well as to improve subsequent performance and relevance of advertisements selected using advertisement selection module 250. Other techniques and features of selecting and presenting advertisements based on voice-based inputs may suitably be employed, as would be apparent.

[025] Referring to Fig. 3, an exemplary method for selecting and presenting advertisements based on a voice-based input is illustrated according to various aspects of the invention. The method may begin in an operation 305, where a voice-based input, including at least a user utterance, may be received at a voice user interface. The voice user interface may include any suitable mechanism for receiving the utterance (e.g., a microphone), and may interface with any suitable voice-enabled device, as would be apparent, including personal navigation devices, personal digital assistants, media devices, telematics devices, personal computers, mobile phones, or others.

[026] Subsequently, one or more requests included in the voice-based input may be identified in an operation 310. For example, the requests may include requests to retrieve information, perform tasks, explore or gather information, or otherwise interact with a system or device. For example, a voice-based input to a navigation device may include a request to calculate a route or retrieve location-based information. In another example, a voice-based input to a mobile phone may include a request to place a telephone call, purchase a ringtone, or record a voice-memo. Furthermore, in various implementations, voice-based inputs may include multiple requests, multi-modal requests, cross-device requests, cross-application requests, or other types of requests. For example, an utterance received in operation 305 may be: "Get me a route to Chang's Restaurant, and call them so I can make a reservation." The utterance may thus include multiple requests, including cross-device requests (e.g., calculate a route using a navigation device, and make a telephone call using a mobile phone), as well as cross-application requests (e.g., search for an address and/or phone number using a voice search engine, and calculate a route using a navigation application).

[027] The requests may be part of a conversational interaction between a user and a system or device, whereby an interpretation of requests in a current utterance may be based upon previous utterances in a current conversation, utterances in previous conversations, context-based information, local and/or global user profiles, or other

Google Exhibit 1019
Page 00605

information. For example, a previous request may be reinterpreted based on information included in subsequent requests, a current request may be interpreted based on information included in previous requests, etc. Furthermore, the conversational interaction may take various forms, including query-based conversations, didactic conversations, exploratory conversations, or other types of conversations. For example, the conversational language processor may identify a type of conversation, and information may be extracted from the utterance accordingly to identify the one or more requests in operation 310. Moreover, the conversational language processor may determine whether any of the requests are incomplete or ambiguous, and action may be taken accordingly (e.g., a system response may prompt a user to clarify an incomplete and/or ambiguous request). The conversational language processor may therefore use various techniques to identify a conversation type, interpret utterances, identify requests, or perform other tasks, such as those described in the aforementioned co-pending patent applications, which are hereby incorporated by reference in their entirety.

[028] Upon identifying the one or more requests, action may be taken based on the identified requests in an operation 315, while one or more advertisements may be selected in an operation 320 (described in greater detail below). For example, one or more context-appropriate applications may be invoked to service the requests in operation 315 (e.g., a voice search engine, a navigation application, an electronic commerce application, or other application may be invoked depending upon the request). Furthermore, in operation 320, information may be communicated to an advertising server to select one or more advertisements related to the request. Thus, as shown in Fig. 3, taking action in operation 315 and selecting advertisements in operation 320 may be related operations (e.g., advertisements may be selected to help in interpreting incomplete and/or ambiguous requests).

[029] Upon taking action in operation 315 (e.g., to service the request) and selecting one or more advertisements in operation 320 (e.g., in relation to the request), an output may be presented to the user in operation 325. The output may indicate a result of the action associated with operation 315. For example, the output may include requested information, an indication of whether a requested task was successfully completed, whether additional information is needed to service the request (e.g., including a prompt for the information), or other information relating to an action based on the request. Furthermore,

Google Exhibit 1019
Page 00606

the output may include advertisements, as selected in operation 320. For example, the output may include text-based, graphic-based, video-based, audio-based, or other types of advertisements, as would be apparent to those skilled in the art. Further, the output may include other types of advertisements, including calls to action (e.g., a location-based coupon or purchase opportunity, trial downloads, or other actionable advertising or marketing).

[030] Advertisements may be selected in relation to a request based on various criteria. For example, an advertisement may be selected based on words or other content of the request, relevant words or content related to the words or content of the request, etc. In another example, the advertisement may be selected based on requested tasks/information (e.g., a request for movie showtimes may result in an advertisement being selected for a particular theater). In yet another example, the advertisement may be selected based on a topic or category associated with the requested tasks/information (e.g., a request to purchase airline tickets may result in an advertisement being selected for a hotel in a destination associated with a reserved flight). In still other examples, the advertisement may be selected based on location information, (e.g., advertisements may be selected based on a proximity to a user geolocation identified using a navigation device), user-specific and/or global user profiles (e.g., advertisements may be selected based on user-specific and/or global preferences, advertiser campaign criteria, etc.).

[031] Content of a voice-based input may be determined based on various criteria, including contextual or conceptual information (e.g., semantic indicators, qualifiers, or other information). For example, a given concept may include various semantically equivalent indicators having an identical meaning. Thus, for instance, a voice-based input may be "Play some tunes!" or "Play some music!" or other variants thereof, each of which may be interpreted as relating to a specific idea (or concept) of "Music." Thus, concept or content information in a request may be used to select an advertisement. For example, a user may request to calculate a route in Seattle, Washington (e.g., "How do I get to the Space Needle?"). Based on a context of the requested task (e.g., "Navigation," "Seattle," etc.), a voice search engine may retrieve an address of the Space Needle and a navigation application may calculate the route. Furthermore, user profile information may indicate that the user is visiting Seattle from out-of-town (e.g., the profile may indicate that the user's home is Sacramento), and therefore, an advertisement for popular points-of-interest

Google Exhibit 1019
Page 00607

in Seattle may be selected. In another example, the user may request information about a sporting event (e.g., "Get me the kickoff time for the Eagles game on Sunday"). Based on a context of the requested information (e.g., "Search," "Sports," "Philadelphia," etc.), the requested information may be retrieved, while an advertisement for Eagles apparel or memorabilia may be selected.

[032] In various instances, concepts, semantic indicators, qualifiers, or other information included in, or inferred from, a request may indicate an exploratory nature for the request. In other words, the exploratory request may identify a goal for a conversation, instead of a particular task to perform or information to retrieve. As such, in various implementations, an advertisement may be selected in operation 320 in an effort to advance the conversation towards the goal. For example, an exploratory request may include a request for a navigation route (e.g., "I feel like going to a museum, find me something interesting"). Based on a context of the requested task (e.g., "Navigation," "Points of Interest," etc.), the goal of the conversation may be identified, and the request may be serviced in operation 315 (e.g., a voice search engine may locate nearby points of interest based on user preferred topics). Further, the advertising application may select an appropriate advertisement in operation 320, where the advertisement may be selected in an attempt to advance the conversation towards the goal. For example, statistical profiles (e.g., user profiles, global profiles, topic-based profiles, etc.) may reflect an affinity between an advertisement for a particular museum and other users sharing similar demographics or other characteristics with the requesting user. Thus, in addition to retrieving information about museums in operation 315, an advertisement for a museum likely to be of interest to the user may be selected in operation 320.

[033] In various instances, a request may include incomplete, ambiguous, unrecognized, or otherwise insufficient semantic indicators, context, qualifiers, or other information needed to identify the request. In other words, the request may include inadequate information to identify or infer a task to perform, information to retrieve, or a goal for a conversation. Thus, as much information as possible may be extracted and/or inferred from the request based on shared knowledge such as context, user or global profile information, previous utterances, previous conversations, etc. As such, servicing the request may include generating a response and/or communicating with an advertising application to advance a conversation toward a serviceable request. For example, servicing the request in operation

Google Exhibit 1019
Page 00608

315 and selecting an advertisement in operation 320 may include generating a response and/or selecting an advertisement to frame a subsequent user input, thereby advancing the conversation.

[034] For example, the request may include incomplete, ambiguous, or unrecognized information (e.g., "Do you know [*mumbled words*] Seattle?"). A context of the requested task may be identified (e.g., "Seattle"), yet the identified context may be insufficient to adequately take action to service the request. Additional information may be inferred based on previous utterances in the conversation, profile information, or other information. However, when the additional information fails to provide adequate information to infer a reasonable hypothesis, servicing the request in operation 315 may include generating a response to frame a subsequent user input and advance the conversation (e.g., information about various topics may be retrieved based on a user's preferred topics). Further, the advertising application may select an advertisement in operation 320 to advance the conversation (e.g., advertisements may be selected based on user and/or global profiles reflecting an affinity between certain advertisements associated with Seattle and user preferences, profiles, etc.). Thus, by selecting an advertisement, indicating dissatisfaction with an advertisement, or otherwise interacting with an advertisement, the interaction may be used to build context and shared knowledge for a subsequent course of the conversation. For example, a user may select an advertisement, and an interpretation of a subsequent voice-based input (e.g., "Call them," "What's the price range?" etc.) may be interpreted with shared knowledge of the advertisement that the voice-based input relates to. Thus, advertisements may be used in a way that enables advertisers to market to consumers, while also improving the consumers' interaction with a device. Other advantages will be apparent to those skilled in the art.

[035] It will be apparent that operation 320 may use various techniques to select advertisements based on voice-based inputs and/or requests included therein. For example, an advertiser may specify a target audience, marketing criteria, campaign strategies, budget constraints, concepts, semantic indicators, related topics, categories, and/or any other suitable information to associate with an advertisement. For instance, advertisers may pay a premium to prioritize an advertisement in relation to similar advertisements (e.g., advertisements associated with competitors). In another example, various statistical profiles may define affinities between advertisements, topics, users, etc.

Google Exhibit 1019
Page 00609

(e.g., based on click-through or conversion rates, or other tracking information, as described in more detail below). Thus, advertisements may be selected in operation 320 using various techniques, including content of the request, an activity/action associated with the request, user profiles, user preferences, statistical metrics, advertiser-specified criteria, to advance a conversation, to resolve ambiguous requests, or in various other ways, as will be apparent.

[036] The output presented to the user in operation 325 may be provided to the user in various ways. For example, in various implementations, the output may include a voice-based or otherwise audible response. In another example, when an associated device includes a display mechanism, the output may be displayed on the display device. It will be apparent that many combinations or variants thereof may be used, such as augmenting a voice-based response with information on a display device. For example, a user may request information about restaurants, and an advertisement may be selected based on a user preference indicating a favorite type of restaurant (e.g., a Chinese restaurant may be selected based on a user profile indicating a preference for Chinese). Therefore, in one example, the output presented in operation 325 may display information about various restaurants matching the requested information, while a voice-based advertisement for the Chinese restaurant may be played to the user (e.g., via a speaker or other suitable mechanism for playing voice back to the user). Many other variations will be apparent (e.g., a graphical advertisement may be displayed on a display device, while a corresponding or different voice-based advertisement may be played audibly).

[037] Subsequent interaction between the user and the presented advertisements may be monitored in a decisional operation 330. For instance, when the user elects to interact with the advertisement, action may be taken based on the interaction in an operation 335. The interaction may take various forms, including additional voice-based inputs or other suitable mechanisms for interacting with advertisements (e.g., clicking on an advertisement displayed on a personal digital assistant using an associated stylus). For example, a user may initially request information from a voice-enabled media device (e.g., a satellite radio player) about a song currently playing (e.g., "What is this song?"). In addition to outputting the requested information about the song (e.g., "This song is Double Barrel by Dave and Ansel Collins."), a selected advertisement may enable the user to purchase a ringtone for a mobile phone that corresponds to the song. In this example, the interaction may include a request to purchase the ringtone (e.g., "Yeah, I'll buy that"), and action taken in operation

Google Exhibit 1019
Page 00610

335 may include completing a transaction for the ringtone and/or downloading the ringtone to the mobile phone. Furthermore, additional advertisements may be selected in an operation 340 based on the interaction, using similar techniques as described in connection with operation 320 (e.g., advertisements for additional ringtones, similar musicians, etc. may be selected). Processing may subsequently return to operation 325 to present output resulting from the interaction.

[038] User advertisement interaction may be tracked in an operation 345. For example, operation 345 may track historical data about users, conversations, topics, contexts, or other criteria to associate information with the selected advertisement. The tracking information may therefore be used to build statistical profiles defining affinities, click-through or conversion rates, or other information about various advertisements, topics, or other criteria on a user-specific and/or a global-user level. Thus, clusters or mappings may be created between advertisements, topics, concepts, demographics, or other criteria based on user behavior with the advertisements (e.g., whether a user interacts with the advertisement in operation 330).

[039] For instance, certain advertisements may experience high click-through rates in relation to a first context and/or topic, but low click-through rates in relation to a second context and/or topic, and therefore, when requests relate to the first context and/or topic, the advertisement may be more likely to be selected in subsequent operations 320/340. In another example, global statistical profiles may indicate that an advertisement experiences more click-throughs by users of a particular demographic, and therefore, the advertisement may be more likely to be selected for users falling within the demographic. Many different techniques for tracking and building statistical profiles will be apparent.

[040] Implementations of the invention may be made in hardware, firmware, software, or any combination thereof. The invention may also be implemented as instructions stored on a machine-readable medium, which may be read and executed by one or more processors. A machine-readable medium may include any mechanism for storing or transmitting information in a form readable by a machine (e.g., a computing device). For example, a machine-readable storage medium may include read only memory, random access memory, magnetic disk storage media, optical storage media, flash memory devices, and others, and a machine-readable transmission media may include forms of propagated signals, such as carrier waves, infrared signals, digital signals, and others. Further, firmware, software,

Google Exhibit 1019
Page 00611

routines, or instructions may be described in the above disclosure in terms of specific exemplary aspects and implementations of the invention, and performing certain actions. However, it will be apparent that such descriptions are merely for convenience and that such actions in fact result from computing devices, processors, controllers, or other devices executing the firmware, software, routines, or instructions.

[041] Aspects and implementations may be described as including a particular feature, structure, or characteristic, but every aspect or implementation may not necessarily include the particular feature, structure, or characteristic. Further, when a particular feature, structure, or characteristic is described in connection with an aspect or implementation, it will be apparent to effect such feature, structure, or characteristic in connection with other aspects or implementations whether or not explicitly described. Thus, various changes and modifications may be made, without departing from the scope and spirit of the invention. The specification and drawings are to be regarded as exemplary only, and the scope of the invention is to be determined solely by the appended claims.

Google Exhibit 1019
Page 00612

CLAIMS

What is claimed is:

1. A method for selecting and presenting advertisements in response to voice-based inputs, the method comprising:

receiving a voice-based input, the received input including a request;

retrieving information from at least one knowledge source based on the received request;

selecting an advertisement based on the received request and the retrieved information; and

presenting the selected advertisement to a user.

2. The method of claim 1, wherein the retrieved information is related to one or more of content of the voice-based input, an action associated with the request, a location of the user, a user-specific profile, or a global user profile.

3. The method of claim 1, wherein selecting the advertisement includes identifying an advertisement related to resolving the request when the received input includes insufficient information to resolve the request.

4. The method of claim 1, further comprising tracking an interaction between the user and the presented advertisement.

5. The method of claim 4, wherein the tracking is used to build statistical profiles of one or more users.

6. The method of claim 5, wherein the statistical profiles identify affinities among one or more advertisements, concepts, categories, and/or user profiles.

7. The method of claim 4, wherein the tracking is used to build a context for interpreting subsequent voice-based inputs.

Google Exhibit 1019
Page 00613

8. The method of claim 4, further comprising taking an action in response to the interaction.

9. The method of claim 8, wherein the action includes performing a task and/or retrieving information based on the interaction.

10. The method of claim 1, wherein the at least one knowledge source includes public information generic to the user, and private information specific to the user.

11. A system for selecting and presenting advertisements in response to voice-based inputs, the system comprising:

at least one input device that receives a voice-based input, the received input including a request; and

at least one processor operable to:

retrieve information from at least one knowledge source based on the received request;

select an advertisement based on the received request and the retrieved information; and

present the selected advertisement to a user.

12. The system of claim 11, wherein the retrieved information is related to one or more of content of the voice-based input, an action associated with the request, a location of the user, a user-specific profile, or a global user profile.

13. The system of claim 11, wherein selecting the advertisement includes identifying an advertisement related to resolving the request when the received input includes insufficient information to resolve the request.

14. The system of claim 11, further comprising tracking an interaction between the user and the presented advertisement.

Google Exhibit 1019
Page 00614

15. The system of claim 14, wherein the tracking is used to build statistical profiles of one or more users.

16. The system of claim 15, wherein the statistical profiles identify affinities among one or more advertisements, concepts, categories, and/or user profiles.

17. The system of claim 14, wherein the tracking is used to build a context for interpreting subsequent voice-based inputs.

18. The system of claim 14, further comprising taking an action in response to the interaction.

19. The system of claim 18, wherein the action includes performing a task and/or retrieving information based on the interaction.

20. The system of claim 11, wherein the at least one knowledge source includes public information generic to the user, and private information specific to the user.

Google Exhibit 1019
Page 00615

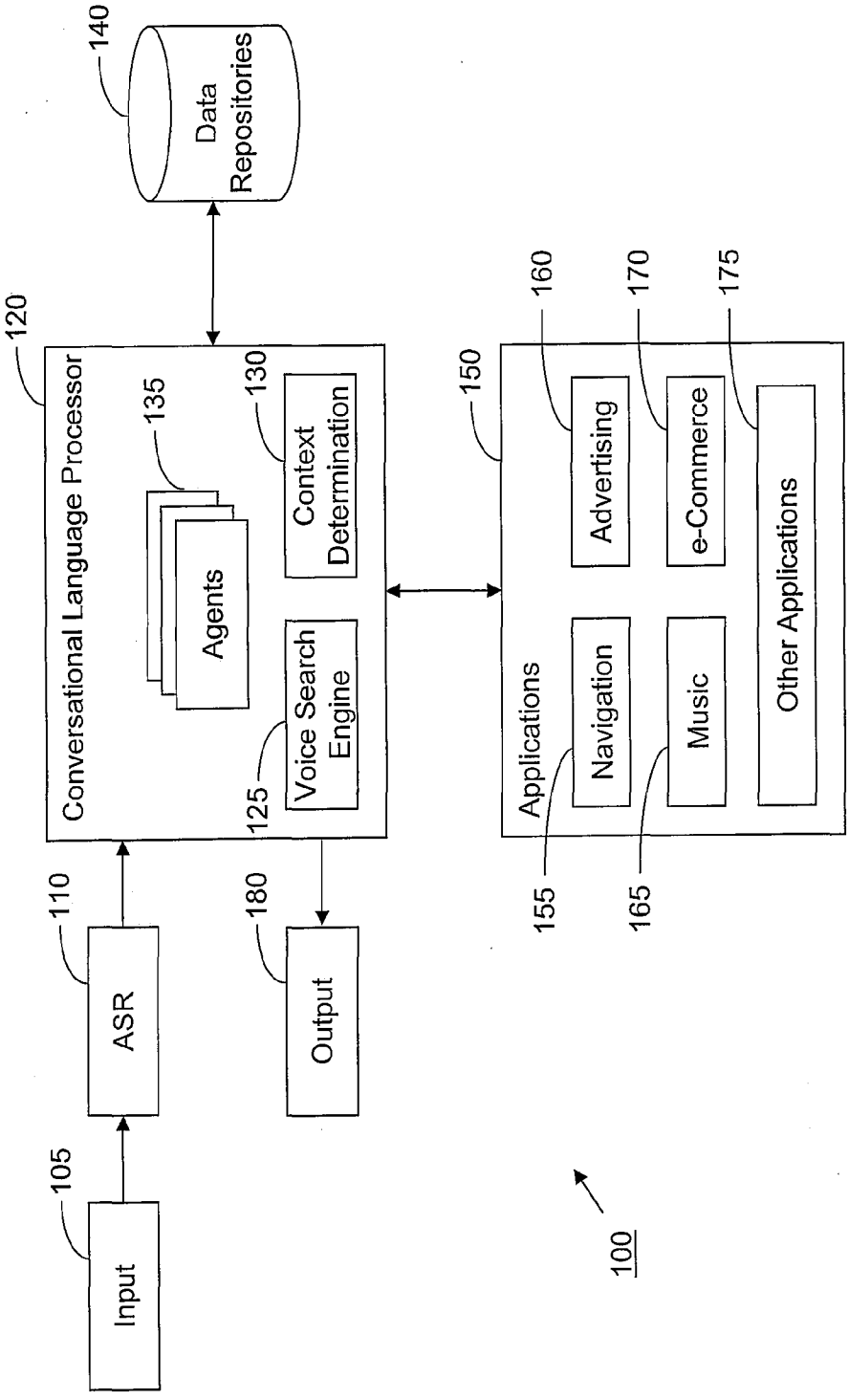

140
Data Repositories
120
Conversational Language Processor
135
Agents
130
Context Determination
125
Voice Search Engine
150
Applications
160
Advertising
170
e-Commerce
175
Other Applications
Navigation
Music
155
165
110
ASR
180
Output
105
Input
100

Fig. 1

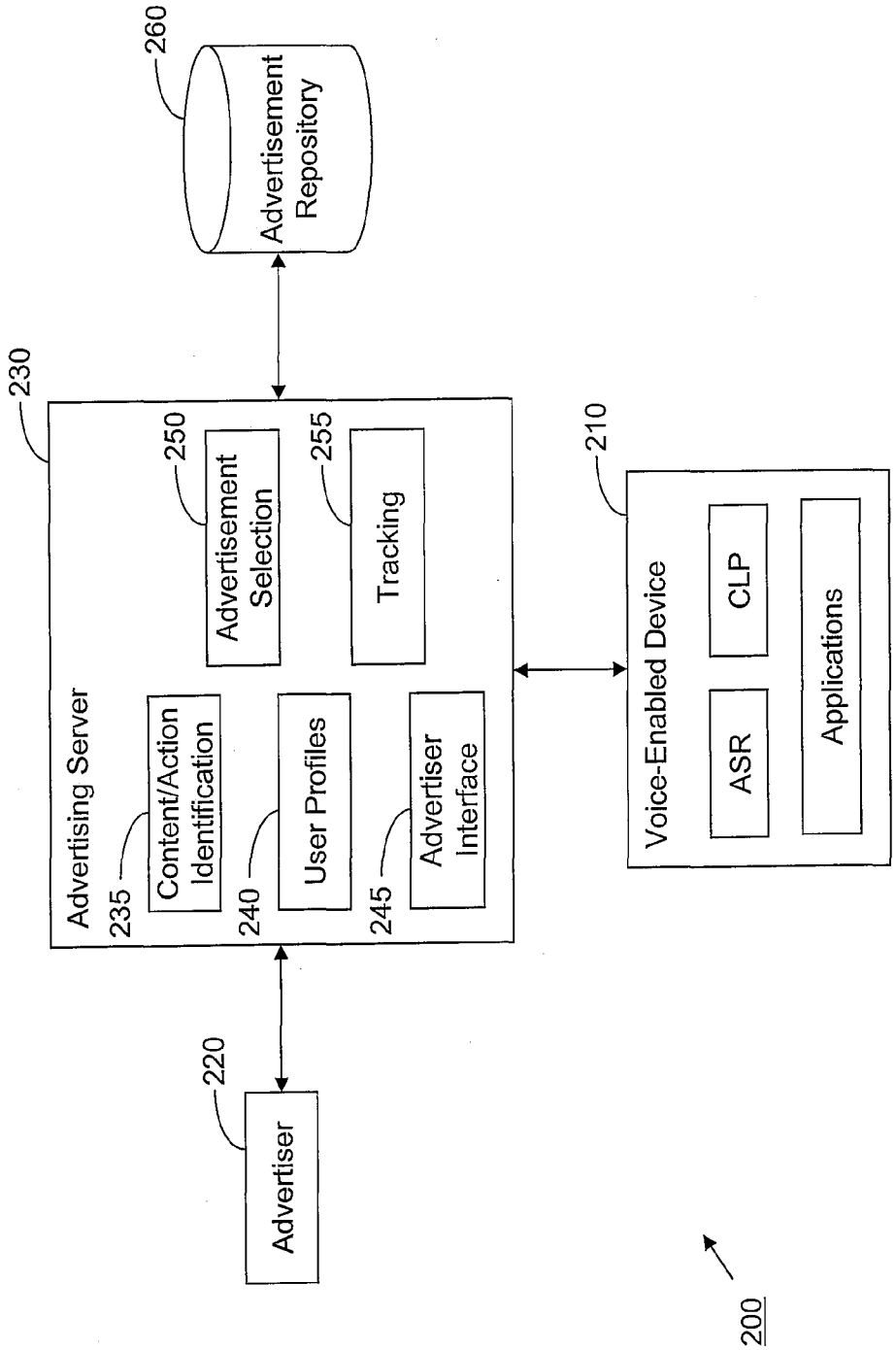

260
Advertisement Repository
230
Advertising Server
250
Advertisement Selection
255
Tracking
235
Content/Action Identification
240
User Profiles
245
Advertiser Interface
210
Voice-Enabled Device
ASR
CLP
Applications
220
Advertiser
200

Fig. 2

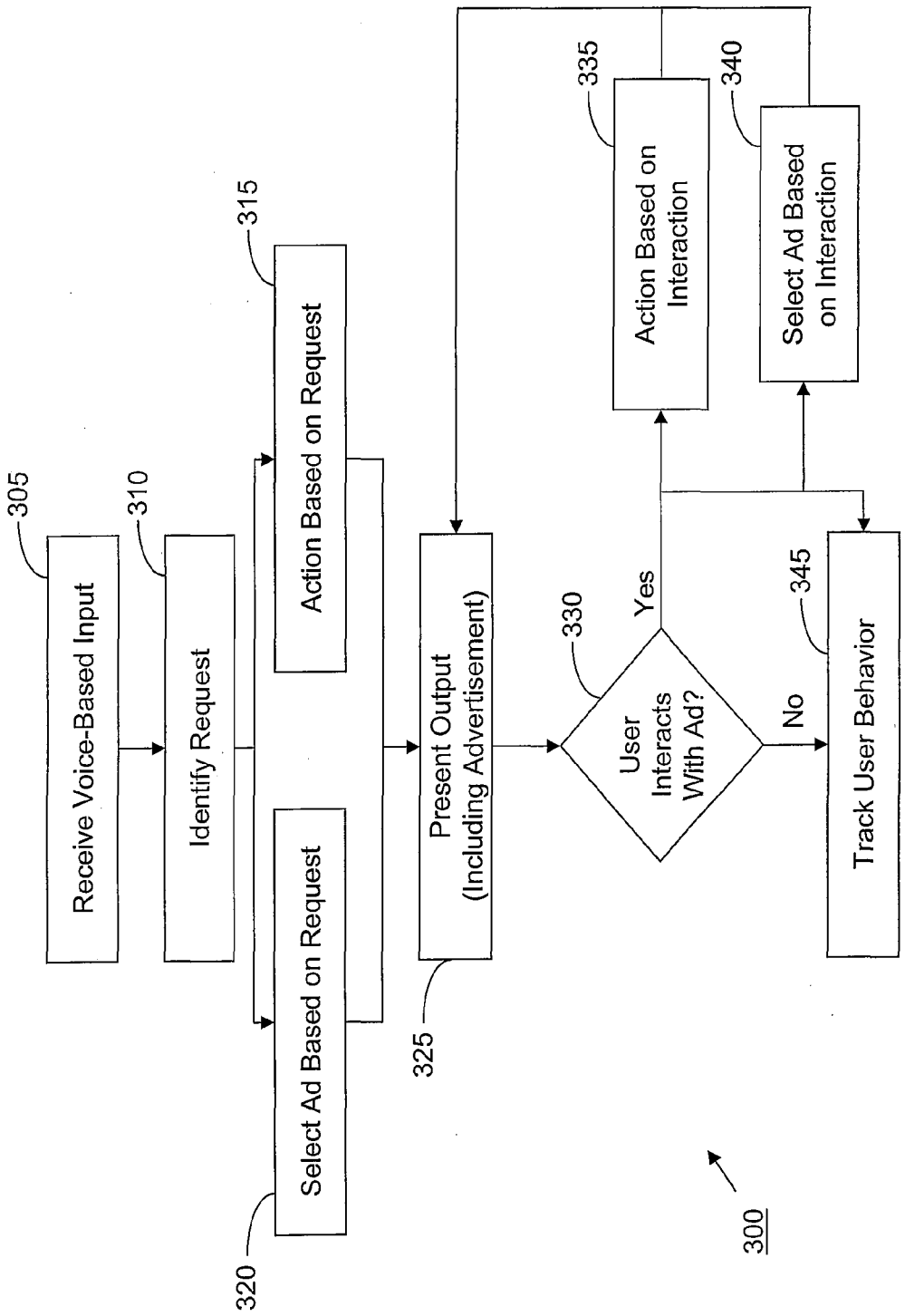

305
Receive Voice-Based Input
310
Identify Request
315
Action Based on Request
320
Select Ad Based on Request
325
Present Output (Including Advertisement)
330
User Interacts With Ad?
Yes
No
335
Action Based on Interaction
340
Select Ad Based on Interaction
345
Track User Behavior
300

Fig. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 08/53157

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06Q 30/00
USPC - 705/14
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - G06Q 30/00 (2008.04)
USPC - 705/14

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
705/1, 7, 8; 725/37, 38, 42

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(USPT,PGPB,EPAB,JPAB); Google Scholar; Google
Search Terms Used: method to provide/select/display advertisement, advertisement, voice input, knowledge source, database, track, profile, user profile, advertisement database/source

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,696,965 A (DEDRICK) 09 December 1997 (09.12.1997) entire document, especially col 4, ln 60-66, col 5, ln 10-20, col 6, ln 12-30 and col 8, ln 24 - col 9, ln 10.	1-20

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 April 2008 (24.04.2008)	22 MAY 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 October 2008 (02.10.2008)

PCT

(10) International Publication Number
WO 2008/118195 A2

(51) International Patent Classification:
G06F 17/28 (2006.01)

(21) International Application Number:
PCT/US2007/081481

(22) International Filing Date: 16 October 2007 (16.10.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/580,926 16 October 2006 (16.10.2006) US

(71) Applicant *(for all designated States except US)*: **VOICEBOX TECHNOLOGIES, INC.** [US/US]; 11980 Ne 24th Street, Bellevue, WA 98005 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BALDWIN, Larry** [US/US]; 25498 S.e. 274th Place, Maple Valley, WA 98038 (US). **FREEMAN, Tom** [US/US]; 6735 83rd Avenue S.e., Mercer Island, WA 98040 (US). **TJALVE, Michael** [DK/US]; 14904 S.e. 47th Court, Bellevue, WA 98006 (US). **EBERSOLD, Blane** [US/US]; 4721 Admiral Way, S.w., Seattle, WA 98116 (US). **WEIDER, Chris** [US/US]; 2618 Grand Avenue, #b-508, Everett, WA 98201 (US).

(74) Agents: **BARUFKA, Jack, S.** et al.; Pillsbury Winthrop Shaw Pittman Llp, P.O. Box 10500, Mclean, VA 22102 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

(54) Title: SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE

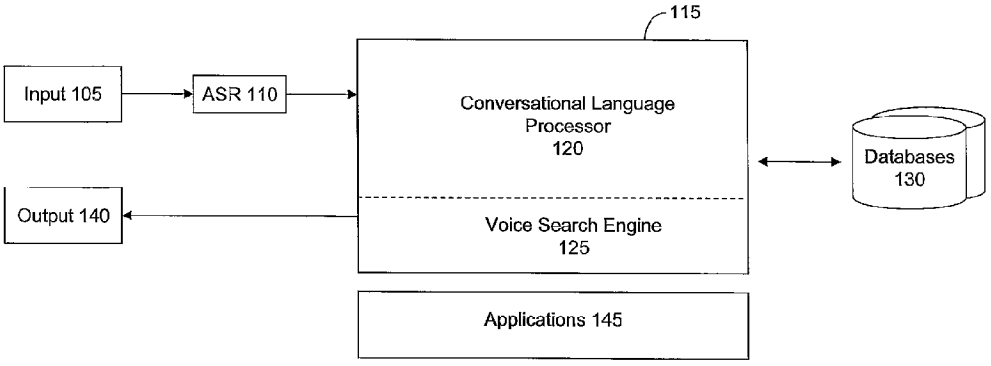


FIG. 1

(57) Abstract: A cooperative conversational voice user interface is provided. The cooperative conversational voice user interface may build upon short-term and long-term shared knowledge to generate one or more explicit and/or implicit hypotheses about an intent of a user utterance. The hypotheses may be ranked based on varying degrees of certainty, and an adaptive response may be generated for the user. Responses may be worded based on the degrees of certainty and to frame an appropriate domain for a subsequent utterance. In one implementation, misrecognitions may be tolerated, and conversational course may be corrected based on subsequent utterances and/or responses.

WO 2008/118195 A2

SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE

FIELD OF THE INVENTION

[001] The invention relates to a cooperative conversational model for a human to machine voice user interface.

BACKGROUND OF THE INVENTION

[002] Advances in technology, particularly within the convergence space, have resulted in an increase in demand for voice recognition software that can exploit technology in ways that are intuitive to humans. While communication between human beings is most often "cooperative," in that information and/or context is shared to advance mutual conversational goals, existing Human-to-Machine interfaces fail to provide the same level of intuitive interaction. For example, each human participant in a conversation can contribute to an exchange for the benefit of the exchange. This is done through shared assumptions and expectations regarding various aspects of the conversation, such as the topic, participant knowledge about the topic, expectations of the other participant's knowledge about the topic, appropriate word usage for the topic and/or participants, conversational development based on previous utterances, the participants' tone or inflection, the quality and quantity of contribution expected from each participant, and many other factors. Participating in conversations that continually build and draw upon shared information is a natural and intuitive way for humans to converse.

[003] In contrast, complex Human-to-Machine interfaces do not allow users to exploit technology in an intuitive way, which inhibits mass-market adoption for various technologies. Incorporating a speech interface helps to alleviate this burden by making interaction easier and faster, but existing speech interfaces (when they actually work) still require significant learning on the part of the user. That is, existing speech interfaces are unable to bridge the gap between archaic Human-to-Machine interfaces and conversational speech that would make interaction with systems feel normal. Users should be able to directly request what they want from a system in a normal, conversational fashion, without having to memorize exact words or phrases. Alternatively, when users are uncertain of particular needs, they should be able to engage the system in a productive, cooperative dialogue to resolve their

requests. Instead, existing speech interfaces force users to dumb down their requests to match simple sets of instructions in simple languages in order to communicate requests in ways that systems can understand. Using existing speech interfaces, there is virtually no option for dialogue between the user and the system to satisfy mutual goals.

[004] Therefore, existing systems lack a conversational speech model that can provide users with the ability to interact with systems in ways that are inherently intuitive to human beings. Existing systems suffer from these and other problems.

SUMMARY OF THE INVENTION

[005] According to various embodiments and aspects of the invention, a cooperative conversational voice user interface may understand free form human utterances, freeing users from being restricted to a fixed set of commands and/or requests. Rather, users can engage in cooperative conversations with a machine to complete a request or series of requests using a natural, intuitive, free form manner of expression.

[006] According to an aspect of the invention, an exemplary system architecture for implementing a cooperative conversational voice user interface is provided. The system may receive an input, which may include a human utterance received by an input device, where the utterance may include one or more requests. As used herein, an "utterance" may be words, syllables, phonemes, or any other audible sound made by a human being. As used herein, a "request" may be a command, directive, or other instruction for a device, computer, or other machine to retrieve information, perform a task, or take some other action. In one implementation, the input may be a multi-modal input, where at least part of the multi-modal input is an utterance. The utterance component of the input may be processed by an Automatic Speech Recognizer to generate one or more preliminary interpretations of the utterance. The one or more preliminary interpretations may then be provided to a conversational speech engine for further processing, where the conversational speech engine may communicate with one or more databases to generate an adaptive conversational response, which may be returned to the user as an output. In one implementation, the output may be a multi-modal output. For example, the utterance may include a request to perform an action, and the output may include a conversational response reporting success or failure, as well as an execution of the action.

[007] According to another aspect of the invention, an exemplary conversational

Google Exhibit 1019
Page 00622

speech engine may generate an adaptive conversational response to a request or series of requests. The conversational speech engine may include a free form voice search module that may understand an utterance made using typical, day-to-day language (i.e., in free form), and may account for variations in how humans normally speak, the vocabulary they use, and the conditions in which they speak. To account for intangible variables of human speech, the free form search module may include models of casual human speech. For example, in one implementation, the free form search module may understand specialized jargon and/or slang, tolerate variations in word order, and tolerate verbalized pauses or stuttered speech. For example, formalized English requests, where a verb precedes a noun, may be treated in an equivalent manner to requests where the noun precedes the verb. In another implementation, compound requests and/or compound tasks with multiple variables may be identified in a single utterance. By identifying all relevant information for completing one or more tasks from a single utterance, advantages may be provided over existing voice user interfaces, such as Command and Control systems that use verbal menus to restrict information that a person can provide at a given point. In another implementation, inferring intended requests from incomplete or ambiguous requests may provide a conversational feel. By modeling what contextual signifiers, qualifiers, or other information may be required to perform a task in an identified context, an adaptive response may be generated, such as prompting a user for missing contextual signifiers, qualifiers, or other information. In one implementation, the response may ask for missing information in a way that most restricts possible interpretations, and the response may be framed to establish a domain for a subsequent user utterance. In another implementation, common alternatives for nouns and verbs may be recognized to reflect variations in usage patterns according to various criteria. Thus, variations in expression may be supported because word order is unimportant or unanticipated, and nouns and/or verbs may be represented in different ways to give simplistic, yet representative, examples. In another implementation, requests may be inferred from contradictory or otherwise inaccurate information, such as when an utterance includes starts and stops, restarts, stutters, run-on sentences, or other imperfect speech. For example, a user may sometimes change their mind, and thus alter the request in mid-utterance, and the imperfect speech feature may nonetheless be able to infer a request based on models of human speech. For example, various models may indicate that a last criterion is most likely to be correct, or intonation, emphasis, stress, use of the word "not," or other

Google Exhibit 1019
Page 00623

models may indicate which criterion is most likely to be correct.

[008] According to another aspect of the invention, the conversational speech engine may include a noise tolerance module that may discard words or noise which has no meaning in a given context to reduce a likelihood of confusion. Moreover, the noise tolerance module may filter out environmental and non-human noise to further reduce a likelihood of confusion. In one implementation, the noise tolerance module may cooperate with other modules and features to filter out words that do not fit into an identified context. For example, the noise tolerance module may filter other human conversations and/or utterances within a range of one or more microphones. For example, a single device may include multiple microphones, or multiple devices may each include one or more microphones, and the noise tolerance module may collate inputs and cooperatively filter out sound by comparing a speech signal from the various microphones. The noise tolerance module may also filter out non-human environmental noise within range of the microphones, out-of-vocabulary words caused by speaker ambiguity or malapropisms, or other noise that may be unrelated to a target request. Performance benchmarks for the noise tolerance module may be defined by noise models based on human criteria. For example, if a driver of a car is 92% likely to be understood by a passenger when traveling at 65 miles-per-hour with windows cracked, then performance benchmarks for the noise tolerance module may have a similar performance under such conditions.

[009] According to another aspect of the invention, the conversational speech engine may include a context determination process that determines one or more contexts for a request to establish meaning within a conversation. The one or more contexts may be determined by having one or more context domain agents compete to determine a most appropriate domain for a given utterance. Once a given domain "wins" the competition, the winning domain may be responsible for establishing or inferring further contexts and updating short-term and long-term shared knowledge. If there is a deadlock between context domain agents, an adaptive conversational response may prompt the user to assist in disambiguating between the deadlocked agents. Moreover, the context determination process may infer intended operations and/or context based on previous utterances and/or requests, whereas existing systems consider each utterance independently, potentially making the same errors over and over again. For example, if a given interpretation turns

Google Exhibit 1019
Page 00624

out to be incorrect, the incorrect interpretation may be removed as a potential interpretation from one or more Automatic Speech Recognizer grammars and/or from possible interpretations determined by the conversational speech engine, thereby assuring that a mistake will not be repeated for an identical utterance.

[010] The context determination process may provide advantages over existing voice user interfaces by continually updating one or more models of an existing context and establishing context as a by-product of a conversation, which cannot be established a priori. Rather, the context determination process may track conversation topics and attempt to fit a current utterance into recent contexts, including switching between contexts as tasks are completed, partially completed, requested, etc. The context determination process may identify one or more context domains for an utterance by defining a collection of related functions that may be useful for users in various context domains. Moreover, each context domain may have relevant vocabularies and thought collections to model word groupings, which when evaluated together, may disambiguate one context domain from another. Thus, eliminating out-of-context words and noise words when searching for relevant combinations may enhance accuracy of inferences. This provides advantages over existing systems that attempt to assign meaning to every component of an utterance (i.e., including out-of-context words and noise words), which results in nearly infinite possible combinations and greater likelihood of confusion. The context determination process may also be self-aware, assigning degrees of certainty to one or more generated hypotheses, where a hypothesis may be developed to account for variations in environmental conditions, speaker ambiguity, accents, or other factors. By identifying a context, capabilities within the context, vocabularies within the context, what tasks are done most often historically in the context, what task was just completed, etc., the context determination process may establish intent from rather meager phonetic clues. Moreover, just as in human-to-human conversation, users may switch contexts at any time without confusion, enabling various context domains to be rapidly selected, without menu-driven dead ends, when an utterance is unambiguous.

[011] According to another aspect of the invention, an exemplary cooperative conversational model may build upon free form voice search, noise tolerance, and context determination to implement a conversational Human-to-Machine interface that reflects human interaction and normal conversational behavior. That is, the cooperative

Google Exhibit 1019
Page 00625

conversational model enables humans and machines to participant in a conversation with an accepted purpose or direction, with each participant contributing to the conversation for the benefit of the conversation. By taking advantage of human presumptions about utterances that humans rely upon, both as speakers and listeners, a Human-to-Machine interface may be analogous to everyday human-to-human conversation. In one implementation, the exemplary cooperative conversation model may take incoming data (shared knowledge) to inform a decision (intelligent hypothesis building), and then may refine the decision and generate a response (adaptive response building).

[012] According to another aspect of the invention, shared knowledge may include both short-term and long-term knowledge. Short-term knowledge may accumulate during a single conversation, where input received during a single conversation may be retained. The shared knowledge may include cross-modality awareness, where in addition to accumulating input relating to user utterances, requests, locations, etc., the shared knowledge may accumulate a current user interface state relating to other modal inputs to further build shared knowledge models. The shared knowledge may be used to build one or more intelligent hypotheses using current and relevant information, build long-term shared knowledge by identifying information with long-term significance, and generate adaptive responses with relevant state and word usage information. Moreover, because cooperative conversations model human conversations, short-term session data may be expired after a psychologically appropriate amount of time, thereby humanizing system behavior, reducing a likelihood of contextual confusion based on stale data, while also adding relevant information from an expired session context to long-term knowledge models. Long-term shared knowledge may generally be user-centric, rather than session-based, where inputs may be accumulated over time to build user, environmental, cognitive, historical, or other long-term knowledge models. Long-term and short-term shared knowledge may be used simultaneously anytime a user engages in a cooperative conversation. Long-term shared knowledge may include explicit and/or implicit user preferences, a history of recent contexts, requests, tasks, etc., user-specific jargon related to vocabularies and/or capabilities of a context, most often used word choices, or other information. The long-term shared knowledge may be used to build one or more intelligent hypotheses using current and relevant information, generate adaptive responses with appropriate word choices when unavailable via short-term shared knowledge, refine long-term shared knowledge models, identify a frequency of

Google Exhibit 1019
Page 00626

specific tasks, identify tasks a user frequently has difficulty with, or provide other information and/or analysis to generate more accurate conversational responses. Shared knowledge may also be used to adapt a level of unprompted support (e.g., for novices versus experienced users, users who are frequently misrecognized, etc.) Thus, shared knowledge may enable a user and a voice user interface to share assumptions and expectations such as topic knowledge, conversation history, word usage, jargon, tone, or other assumptions and/or expectations that facilitate a cooperative conversation between human users and a system.

[013] According to another aspect of the invention, a conversation type may be identified for any given utterance. Categorizing and developing conceptual models for various types of exchanges may consistently align user expectations and domain capabilities. One or more intelligent hypotheses may be generated as to a conversation type by considering conversational goals, participant roles, and/or an allocation of information among the participants. Based on the conversational goals, participant roles, and allocation of information, the intelligent hypotheses may consider various factors to classify a conversation (or utterance) into general types of conversations that can interact with one another to form many more variations and permutations of conversation types (e.g., a conversation type may change dynamically as information is reallocated from one participant to another, or as conversational goals change based on the reallocation of information).

[014] According to another aspect of the invention, the intelligent hypotheses may include one or more hypotheses of a user's intent in an utterance. In addition, the intelligent hypotheses may use short-term and/or long-term shared knowledge to proactively build and evaluate interaction with a user as a conversation progresses or over time. The hypotheses may model human-to-human interaction to include a varying degree of certainty for each hypothesis. That is, just as humans rely on knowledge shared by participants to examine how much and what kind of information was available, the intelligent hypotheses may leverage the identified conversation type and shared knowledge to generate a degree of certainty for each hypothesis.

[015] According to another aspect of the invention, syntactically, grammatically, and contextually sensitive "intelligent responses" may be generated from the intelligent hypotheses that can be used to generate a conversational experience for a user, while

Google Exhibit 1019
Page 00627

also guiding the user to reply in a manner favorable for recognition. The intelligent responses may create a conversational feel by adapting to a user's manner of speaking, framing responses appropriately, and having natural variation and/or personality (e.g., by varying tone, pace, timing, inflection, word use, jargon, and other variables in a verbal or audible response).

[016] According to another aspect of the invention, the intelligent responses may adapt to a user's manner of speaking by using contextual signifiers and grammatical rules to generate one or more sentences that may cooperate with the user. By taking advantage of shared knowledge about how a user utters a request, the responses may be modeled using similar techniques used to recognize requests. The intelligent responses may rate possible responses statistically and/or randomize responses, which creates an opportunity to build an exchange with natural variation and conversational feel. This provides advantages over existing voice user interfaces where input and output is incongruous, as the input is "conversational" and the output is "computerese."

[017] According to another aspect of the invention, the intelligent responses may frame responses to influence a user reply utterance for easy recognition. For example, the responses may be modeled to illicit utterances from the user that may be more likely to result in a completed request. Thus, the responses may conform to a cooperative nature of human dialog and a natural human tendency to "parrot" what was just heard as part of a next utterance. Moreover, knowledge of current context may enhance responses to generate more meaningful conversational responses. Framing the responses may also deal with misrecognitions according to human models. For example, humans frequently remember a number of recent utterances, especially when one or more previous utterances were misrecognized or unrecognized. Another participant in the conversation may limit correction to a part of the utterance that was misrecognized or unrecognized, or over subsequent utterances and/or other interactions, clues may be provided to indicate the initial interpretation was incorrect. Thus, by storing and analyzing multiple utterances, utterances from earlier in a conversation may be corrected as the conversation progresses.

[018] According to another aspect of the invention, the intelligent responses may include multi-modal, or cross-modal, responses to a user. In one implementation, responses may be aware of and control one or more devices and/or interfaces, and users may respond by using whichever input method, or combination of input methods, is most

Google Exhibit 1019
Page 00628

convenient.

[019] According to another aspect of the invention, the intelligent responses may correct a course of a conversation without interrupting conversational flow. That is, even though the intelligent responses may be reasonably "sure," the intelligent responses may nonetheless sometimes be incorrect. While existing voice user interfaces tend to fail on average conversational missteps, normal human interactions may expect missteps and deal with them appropriately. Thus, responses after misrecognitions may be modeled after clarifications, rather than errors, and words may chosen in subsequent responses to move conversation forward and establish an appropriate domain to be explored with the user.

[020] Other objects and advantages of the invention will be apparent to those skilled in the art based on the following drawings and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[021] Fig. 1 is an exemplary block diagram of a system architecture according to one aspect of the invention.

[022] Fig. 2 is an exemplary block diagram of a conversational speech engine according to one aspect of the invention.

[023] Fig. 3 is an exemplary block diagram of a cooperative conversational model according to one aspect of the invention.

DETAILED DESCRIPTION

[024] Referring to Fig. 1, an exemplary system architecture for implementing a cooperative conversational voice user interface is illustrated according to one aspect of the invention. The system may receive an input 105 from a user, where in one implementation, input 105 may be an utterance received by an input device (e.g., a microphone), where the utterance may include one or more requests. Input 105 may also be a multi-modal input, where at least part of the multi-modal input is an utterance. For example, the input device may include a combination of a microphone and a touch-screen device, and input 105 may include an utterance that includes a request relating to a portion of a display on the touch-screen device that the user is touching. For instance, the touch-screen device may be a navigation device, and input 105 may include an utterance

Google Exhibit 1019
Page 00629

of "Give me directions to here," where the user may be requesting directions to a desired destination on the display of the navigation device.

[025] The utterance component of input 105 may be processed by an Automatic Speech Recognizer 110 to generate one or more preliminary interpretations of the utterance. Automatic Speech Recognizer 110 may process the utterance using any suitable technique known in the art. For example, in one implementation, Automatic Speech Recognizer 110 may interpret the utterance using techniques of phonetic dictation to recognize a phoneme stream, as described in copending U.S. Patent Application Ser. No. 11/513,269, entitled "Dynamic Speech Sharpening," which is hereby incorporated by reference in its entirety. The one or more preliminary interpretations generated by Automatic Speech Recognizer 110 may then be provided to a conversational speech engine 115 for further processing. Conversational speech engine 115 may include a conversational language processor 120 and/or a voice search engine 125, described in greater detail in Fig. 2 below. Conversational speech engine 115 may communicate with one or more databases 130 to generate an adaptive conversational response, which may be returned to the user as an output 140. In one implementation, output 140 may be a multi-modal output and/or an interaction with one or more applications 145 to complete the request. For example, output 140 may include a combination of an audible response and a display of a route on a navigation device. For example, the utterance may include a request to perform an action, and output 140 may include a conversational response reporting success or failure, as well as an execution of the action. In addition, in various implementations, Automatic Speech Recognizer 110, conversational speech engine 115, and/or databases 130 may reside locally (e.g., on a user device), remotely (e.g., on a server), or a hybrid model of local and remote processing may be used (e.g., lightweight applications may be processed locally while computationally intensive applications may be processed remotely).

[026] Referring to Fig. 2, an exemplary block diagram is provided illustrating a conversational speech engine 215 according to one aspect of the invention. Conversational speech engine 215 may include a conversational language processor 220 that generates an adaptive conversational response to a request or series of requests using a free form voice search module 245, a noise tolerance module 250, and/or a context determination process 255. According to one aspect of the invention, modules 245-255 may communicate with a voice search engine 225 that includes one or more context

Google Exhibit 1019
Page 00630

domain agents 230 and/or one or more vocabularies 235 to aid in interpreting utterances and generating responses, as described in "Enhancing the VUE™ (Voce-User-Experience) Through Conversational Speech," by Tom Freeman and Larry Baldwin, which is herein incorporated by reference in its entirety. Conversational speech engine 215 may generate an adaptive conversational response to one or more requests, where the requests may depend on unspoken assumptions, incomplete information, context established by previous utterances, user profiles, historical profiles, environmental profiles, or other information. Moreover, conversational speech engine 215 may track which requests have been completed, which requests are being processed, and/or which requests cannot be processed due to incomplete or inaccurate information, and the response may be generated accordingly.

[027] According to one aspect of the invention, free form voice search module 245 may understand an utterance made using typical, day-to-day language (i.e., in free form), and may account for variations in how humans normally speak, the vocabulary they use, and the conditions in which they speak. Because variables such as stress, distraction, and serendipity are always different and infinitely varied, free form search module 245 may be designed with a goal of understanding that no human will come to the same Human-to-Machine interface situation in the same way twice. Thus, free form search module 245 may implement one or more features that model casual human speech. In various implementations, free form search module 245 may include, among other things, a free form utterance feature, a one-step access feature, an inferencing intended operations feature, an alternative expression feature, and/or an imperfect speech feature.

[028] The free form utterance feature may understand specialized jargon and/or slang, tolerate variations in word order (e.g., whether a subject of a request comes before or after a verb may be irrelevant), and tolerate verbalized pauses (e.g., "um," "ah," "eh," and other utterances without meaning). For example, the free form utterance feature may treat formalized English verb-before-noun requests in an equivalent manner to free form requests where a noun may precede a verb. For example, user utterances of "Change it to the Squizz" and "You know, um, that Squizz channel, ah, switch it there" may be treated equivalently (where Squizz is a channel on XM Satellite Radio). In either case, the free form utterance feature is able to identify "Squizz" as a subject of the utterance and "Change it" or "switch it" as a verb or request for the utterance (e.g., by

Google Exhibit 1019
Page 00631

cooperating with context determination process 255, or other features, and identifying a relevant context domain agent 230 and/or vocabulary 235 to interpret the utterance).

[029] The one-step access feature may understand utterances that include compound requests with multiple variables. For example, a user utterance may be "What is the forecast for Boston this weekend?" The one-step access feature may identify "weather" as a context (e.g., by cooperating with context determination process 255, or other features, and identifying "forecast" as a synonym of "weather"), and search for a city equal to "Boston" and a time equal to "weekend." By identifying all relevant information for completing a task from a single utterance, the one-step access feature may overcome drawbacks of existing voice user interfaces, such as Command and Control systems that use verbal menus to restrict information that a person can provide at a given point (e.g., a Command and Control system for a phone directory service may say: "State please," . . . "City please," . . . "What listing," etc.). Moreover, some utterances may include compound requests, and the one-step access feature may decompose the compound requests into sub-tasks. For example, a user utterance of "I need to be at a meeting tomorrow in San Francisco at 8:00 am" may be decomposed into a set of sub-tasks such as (1) checking availability and reserving a flight on an evening before the meeting, (2) checking availability and reserving a hotel, (3) checking availability and reserving a car, etc., where users may further designate preferences for various tasks (e.g., first check availability on an airline for which the user is a frequent flyer). Depending on a level of shared knowledge about a user's preferences and/or historical patterns, the one-step access feature may infer additional tasks from a request. For example, in the above example, the one-step access feature may also check a weather forecast, and if the weather is "nice" (as defined by the user preferences and/or as inferred from historical patterns), the one-step access feature may schedule a tee-time at a preferred golf course in San Francisco.

[030] The inferencing intended operations feature may identify an intended request from incomplete or ambiguous requests. For example, when a user utters "Route <indecipherable> Chicago <indecipherable> here," where the user intended to say "Route calculation to Chicago from here," the inferencing intended operations feature may model what is required to calculate a route (an origination point and a destination point). Because the utterance includes the origination point and the destination point, a request to calculate a route from the user's present location to Chicago may be inferred.

Google Exhibit 1019
Page 00632

Similarly, when the inferencing intended operations feature does not have sufficient information to infer a complete request, an adaptive conversational response may be generated to prompt the user for missing information. For example, when an utterance includes a request for a stock quote but not a company name (e.g., "Get me the stock price for <indecipherable>"), the response may be "What company's stock quote do you want?" The user may then provide an utterance including the company name, and the request may be completed. In one implementation, the response may ask for missing information in a way that most restricts possible interpretations (e.g., in a request for a task that requires both a city and a state, the state may be asked for first because there are fewer states than cities). Moreover, the inferencing intended operations feature may model compound tasks and/or requests by maintaining context and identifying relevant and/or missing information at both a composite and sub-task level.

[031] The alternative expression feature may recognize common alternatives for nouns and verbs to reflect variations in usage patterns according to various criteria. For example, users may vary expression based on age, socio-economics, ethnicity, user whims, or other factors. Thus, the alternative expression feature may support variations in expression where word order is unimportant or unanticipated. Alternatives in expression based on various criteria or demographics may be loaded into context domain agents 230 and/or vocabularies 235, and the alternative expression feature may update context domain agents 230 and/or vocabularies 235 based on inferred or newly discovered variations. In one implementation, conversational speech engine 215 may include a subscription interface to update changes to context domain agents 230 and/or vocabularies 235 (e.g., a repository may aggregate various user utterances and deploy updates system wide). In operation, the alternative expression feature may allow nouns and/or verbs to be represented in different ways to give simplistic, yet representative, examples. For example, a user interested in a weather forecast for Washington, D.C. may provide any of the following utterances, each of which are interpreted equivalently: "What's the weather like in DC," "Is it raining inside the Beltway," Gimme the forecast for the capital," etc. Similarly, utterances of "Go to my home," "Go home," "Show route to home," and "I would like to know my way home" may all be interpreted equivalently, where a user profile may include the user's home address and a navigation route to the home address may be calculated.

[032] The imperfect speech feature may be able to infer requests from contradictory or

Google Exhibit 1019
Page 00633

otherwise inaccurate information, such as when an utterance includes starts and stops, restarts, stutters, run-on sentences, or other imperfect speech. For example, a user may sometimes change their mind, and thus alter the request in mid-utterance, and the imperfect speech feature may nonetheless be able to infer a request based on models of human speech. For example, for an utterance of "Well, I wanna . . . Mexi . . . no, steak restaurant please, I'm hungry," existing voice user interfaces make no assumptions regarding models of human speech and would be unable to infer whether the user wanted a Mexican or steak restaurant. The imperfect speech feature overcomes these drawbacks by using various models of human understanding that may indicate that a last criterion is most likely to be correct, or intonation, emphasis, stress, use of the word "not," or other models may indicate which criterion is most likely to be correct. Thus, in the above example, the imperfect speech feature may infer that the user wants a steak restaurant.

[033] According to one aspect of the invention, noise tolerance module 250 may be closely related to the imperfect speech feature, and may operate to discard words or noise that has no meaning in a given context so as not to create confusion. Moreover, noise tolerance module 250 may filter out environmental and non-human noise to further reduce a likelihood of confusion. In one implementation, noise tolerance module 250 may cooperate with other modules and features to filter out words that do not fit into a context. For example, one or more contexts may be identified, and words that have no meaning with respect to system capabilities, random human utterances without meaning and other noise may be filtered out. Thus, noise tolerance module 250 may model real-world conditions to identify meaningful requests. For example, noise tolerance module 250 may filter other human conversations and/or utterances within a range of one or more microphones, For example, a single device may include multiple microphones, or multiple devices may each include one or more microphones, and the noise tolerance module may collate inputs and cooperatively filter out sound by comparing a speech signal from the various microphones. Noise tolerance module 250 may also filter out non-human environmental noise within the range of the microphones, out-of-vocabulary words, which could be a result of speaker ambiguity or malapropisms, or other noise that may be unrelated to a target request. Noise models in noise tolerance module 250 may define performance benchmarks based on human criteria. For example, if a driver of a car, traveling at 65 miles-per-hour, with windows cracked is 92% likely to be understood by a passenger, then noise tolerance module 250 may have a similar performance under

Google Exhibit 1019
Page 00634

those conditions.

[034] According to one aspect of the invention, conversational speech engine 215 may include a context determination process 255 that determines one or more contexts for a request to establish meaning within a conversation. The one or more contexts may be determined by having one or more context domain agents compete to determine a most appropriate domain for a given utterance, as described in copending U.S. Patent Application Ser. No. 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," and copending U.S. Patent Application Ser. No. 11/212,693, entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," both of which are hereby incorporated by reference in their entirety. Once a given context domain agent "wins" the competition, the winning agent may be responsible for establishing or inferring further contexts and updating short-term and long-term shared knowledge. If there is a deadlock between context domain agents, an adaptive conversational response may prompt the user to assist in disambiguating between the deadlocked agents. For example, a user utterance of "What about traffic?" may have a distinct meaning in various contexts. That is, "traffic" may have a first meaning when the user is querying a system's media player (i.e., "traffic" would be a Rock and Roll band led by singer/songwriter Steve Winwood), a second meaning when the user is querying a search interface regarding Michael Douglas films (i.e., "traffic" would be a film directed by Steven Soderbergh), a third meaning when the user is querying a navigation device for directions to an airport (i.e., "traffic" would be related to conditions on roads along a route to the airport).

[035] Moreover, context determination process 255 may infer intended operations and/or context based on previous utterances and/or requests, whereas existing systems consider each utterance independently, potentially making the same errors over and over again. For example, if a given interpretation turns out to be incorrect, the incorrect interpretation may be removed as a potential interpretation from one or more Automatic Speech Recognizer grammars and/or from possible subsequent interpretations determined by context determination process 255, thereby assuring that a mistake will not be repeated for an identical utterance.

[036] Context determination process 255 may overcome drawbacks of existing systems by continually updating one or more models of an existing context, where establishing

Google Exhibit 1019
Page 00635

context may be a by-product of a conversation, which cannot be established a priori. Context determination process 255 may establish a first context domain, change to a second context domain, change back to the first context domain, and so on, as tasks are completed, partially completed, requested, etc, and a context stack may track conversation topics and attempt to fit a current utterance into a most-recent context, next-most-recent topic, etc., traversing the context stack until a most likely intent can be established. For example, a user may utter "What's the traffic report," and context determination process 255 may establish Traffic as a context, and return an output including a traffic report, which does not happen to mention traffic on Interstate-5. The user may then utter "What about I-5?" and context determination process 255 may know that the current context is Traffic, a traffic report including information about Interstate-5 may be searched for, and the traffic report indicating that Interstate-5 is crowded may be returned as an output. The user may then utter "Is there a faster way?" and context determination module 255 may know that the current context is still Traffic, and may search for routes to a specified destination with light traffic and avoiding Interstate-5. Moreover, context determination process 255 may build context based on user profiles, environmental profiles, historical profiles, or other information to further refine the context. For example, the profiles may indicate that Interstate-5 is a typical route taken Monday through Friday.

[037] The profiles may be particularly meaningful when attempting to disambiguate between contexts where a word has different meanings in different contexts. For example, a user may utter "What's the weather in Seattle?" and context determination process 255 may establish Weather as a context, as well as establishing Seattle as an environmental context. The user may then utter "and Portland?" and context determination process 255 may return a weather report for Portland, Oregon based on the Weather and an environmental proximity between Portland, Oregon and Seattle, Washington. The user may then ask "What time does the game start?"" and a search for sports events with teams from Seattle and/or Portland may occur, with results presented conversationally according to methods described in greater detail below in Fig. 3. Correlatively, had user originally uttered "What's the weather in Portsmouth, New Hampshire," in the second utterance, context determination process 255 may instead retrieve a weather report for Portland, Maine based on an environmental proximity to New Hampshire. Moreover, when environmental profiles, contextual shared knowledge, and/or other short-term and/or long-term shared knowledge does not provide enough information to

Google Exhibit 1019
Page 00636

disambiguate between possibilities, responses may prompt the user with a request for further information (e.g., "Did you mean Portland, Maine, or Portland, Oregon?").

[038] Context determination process 255 may cooperate with context domain agents 230, where each context domain agent 230 may define a collection of related functions that may be useful for users. Moreover, each context domain agent 230 may include a relevant vocabulary 235 and thought collections that model word groupings, which when evaluated together, may disambiguate one context domain from another (e.g., a Music context domain agent 230 may include a vocabulary 235 for songs, artists, albums, etc., whereas a Stock context domain agent 230 may include a vocabulary 235 for company names, ticker symbols, financial metrics, etc.). Thus, accuracy in identifying meaning may be enhanced by eliminating out-of-context words and noise words when searching for relevant combinations. In contrast, existing systems attempt to assign meaning to every component of an utterance (e.g., including out-of-context words and noise words), which results in nearly infinite possible combinations and greater likelihood of confusion. Moreover, context domain agents 230 may include metadata for each criteria to further assist in interpreting utterances, inferring intent, completing incomplete requests, etc. (e.g., a Space Needle vocabulary word may include metadata for Seattle, landmark, tourism, Sky City restaurant, etc.). Given a disambiguated criterion, context determination process 255 may thus be able to automatically determine other information needed to complete a request, discard importance of word order, and perform other enhancements for conversational speech.

[039] Context domain agents 230 may also be self-aware, assigning degrees of certainty to one or more generated hypotheses, where a hypothesis may be developed to account for variations in environmental conditions, speaker ambiguity, accents, or other factors. Conceptually, context domain agents 230 may be designed to model utterances as a hard-of-hearing person would at a noisy party. By identifying a context, capabilities within the context, vocabularies within the context, what tasks are done most often historically in the context, what task was just completed, etc., a context domain agent 230 may establish intent from rather meager phonetic clues. Moreover, the context stack may be one of a plurality of components for establishing context, and thus not a constraint upon the user. All context domains may be accessible, allowing the user to switch contexts at any time without confusion. Thus, just as in human- to-human conversation, context domains

Google Exhibit 1019
Page 00637

may be rapidly selected, without menu-driven dead ends, when an utterance is unambiguous. For example, a user may utter, "Please call Rich Kennewick on his cell phone," and a system response of "Do you wish me to call Rich Kennewick on his cell?" may be generated. The user may decide at that point to call Rich Kennewick later, and instead, listen to some music. Thus, the user may then utter, "No, play the Louis Armstrong version of Body and Soul from my iPod," and a system response of "Playing Body and Soul by Louis Armstrong" may be generated as Body and Soul is played through a media player. In this example, the later utterance has no contextual connection to the first utterance, yet because request criteria in the utterances are unambiguous, contexts can be switched easily without relying on the context stack.

[040] Referring to Fig. 3, an exemplary cooperative conversational model 300 is illustrated according to an aspect of the invention. Cooperative conversational model 300 may build upon free form voice search 245, noise tolerance 250, and context determination 255 to implement a conversational Human-to-Machine interface that reflects how humans interact with each other and their normal behavior in conversation. Simply put, cooperative conversational model 300 enables humans and machines to participate in a conversation with an accepted purpose or direction, with each participant contributing to the conversation for the benefit of the conversation. That is, cooperative conversational model 300 incorporates technology and process-flow that takes advantage of human presumptions about utterances that humans rely upon, both as speakers and listeners, thereby creating a Human-to-Machine interface that is analogous to everyday human-to-human conversation. In one implementation, a cooperative conversation may take incoming data (shared knowledge) 305 to inform a decision (intelligent hypothesis building) 310, and then may refine the decision and generate a response (adaptive response building) 315.

[041] According to one aspect of the invention, shared knowledge 305 includes both short-term and long-term knowledge about incoming data. Short-term knowledge may accumulate during a single conversation, while long-term knowledge may accumulate over time to build user profiles, environmental profiles, historical profiles, cognitive profiles, etc.

[042] Input received during a single conversation may be retained in a Session Input Accumulator. The Session Input Accumulator may include cross-modality

Google Exhibit 1019
Page 00638

awareness, where in addition to accumulating input relating to user utterances, requests, locations, etc., the Session Input Accumulator may accumulate a current user interface state relating to other modal inputs to further build shared knowledge models and more accurate adaptive responses (e.g., when a user utters a request relating to a portion of a touch-screen device, as described above). For example, the Session Input Accumulator may accumulate inputs including recognition text for each utterance, a recorded speech file for each utterance, a list-item selection history, a graphical user interface manipulation history, or other input data. Thus, the Session Input Accumulator may populate Intelligent Hypothesis Builder 310 with current and relevant information, build long-term shared knowledge by identifying information with long-term significance, provide Adaptive Response Builder 315 with relevant state and word usage information, retain recent contexts for use with Intelligent Hypothesis Builder 310, and/or retain utterances for reprocessing during multi-pass evaluations. Moreover, because cooperative conversations 300 model human conversations, short-term session data may be expired after a psychologically appropriate amount of time, thereby humanizing system behavior. For example, a human is unlikely to recall a context of a conversation from two years ago, but because the context would be identifiable by a machine, session context is expired after a predetermined amount of time to reduce a likelihood of contextual confusion based on stale data. However, relevant information from an expired session context may nonetheless be added to user, historical, environmental, cognitive, or other long-term knowledge models.

[043] Long-term shared knowledge may generally be user-centric, rather than session-based. That is, inputs may be accumulated over time to build user, environmental, cognitive, historical, or other long-term knowledge models. Long-term and short-term shared knowledge (collectively, shared knowledge 305) may be used simultaneously anytime a user engages in a cooperative conversation 300. Long-term shared knowledge may include explicit and/or implicit user preferences, a history of most recently used agents, contexts, requests, tasks, etc., user-specific jargon related to vocabularies and/or capabilities of an agent and/or context, most often used word choices, or other information. The long-term shared knowledge may be used to populate Intelligent Hypothesis Builder 310 with current and relevant information, provide Adaptive Response Builder 315 with appropriate word choices when the appropriate word choices are unavailable via the Session Input Accumulator, refine long-term shared

Google Exhibit 1019
Page 00639

knowledge models, identify a frequency of specific tasks, identify tasks a user frequently has difficulty with, or provide other information and/or analysis to generate more accurate conversational responses.

[044] As described above, shared knowledge 305 may be used to populate Intelligent Hypothesis Builder 310, such that a user and a voice user interface may share assumptions and expectations such as topic knowledge, conversation history, word usage, jargon, tone (e.g., formal, humorous, terse, etc.), or other assumptions and/or expectations that facilitate interaction at a Human-to-Machine interface.

[045] According to an aspect of the invention, one component of a successful cooperative conversation may be identifying a type of conversation from an utterance. By categorizing and developing conceptual models for various types of exchanges, user expectations and domain capabilities may be consistently aligned. Intelligent Hypothesis Builder 310 may generate a hypothesis as to a conversation type by considering conversational goals, participant roles, and/or an allocation of information among the participants. Conversational goals may broadly include: (1) getting a discrete piece of information or performing a discrete task, (2) gathering related pieces of information to make a decision, and/or (3) disseminating or gathering large amounts of information to build expertise. Participant roles may broadly include: (1) a leader that controls a conversation, (2) a supporter that follows the leader and provides input as requested, and/or (3) a consumer that uses information. Information may be held by one or more of the participants at the outset of a conversation, where a participant may hold most (or all) of the information, little (or none) of the information, or the information may be allocated roughly equally amongst the participants. Based on the conversational goals, participant roles, and allocation of information, Intelligent Hypothesis Builder 310 may consider various factors to classify a conversation (or utterance) into general types of conversations that can interact with one another to form many more variations and permutations of conversation types (e.g., a conversation type may change dynamically as information is reallocated from one participant to another, or as conversational goals change based on the reallocation of information).

[046] For example, in one implementation, a query conversation may include a conversational goal of getting a discrete piece of information or performing a particular task, where a leader of the query conversation may have a specific goal in

Google Exhibit 1019
Page 00640

mind and may lead the conversation toward achieving the goal. The other participant may hold the information and may support the leader by providing the information. In a didactic conversation, a leader of the conversation may control information desired by a supporter of the conversation. The supporter's role may be limited to regulating an overall progression of the conversation and interjecting queries for clarification. In an exploratory conversation, both participants share leader and supporter roles, and the conversation may have no specific goal, or the goal may be improvised as the conversation progresses. Based on this model, Intelligent Hypothesis Builder 310 may broadly categorize a conversation (or utterance) according to the following diagram:

QUERY		
	Participant A	*Participant B*
	User	Voice User Interface
GOAL	Get information/action	Provide information/action
ROLE	Leader/Consumer	Supporter/Dispenser
INFORMATION ALLOCATION	Less	More

Google Exhibit 1019
Page 00641

DIDACTIC		
	Participant A	*Participant B*
	User	Voice User Interface
GOAL	Get information	Provide information
ROLE	Follower/Consumer	Leader/Dispenser
INFORMATION ALLOCATION	Less	More

EXPLORATORY		
	Participant A	*Participant B*
	User	Voice User Interface
GOAL	Gather/share information	Gather/share information
ROLE	Follower/Consumer and Leader/Dispenser	Follower/Consumer and Leader/Dispenser
INFORMATION ALLOCATION	Equal or alternating	Equal or alternating

[047] Intelligent Hypothesis Builder 310 may use an identified conversation type to assist in generating a set of hypotheses as to a user's intent in an utterance. In addition, Intelligent Hypothesis Builder 310 may use short-term shared knowledge from the Session Input Accumulator to proactively build and evaluate interaction with a user as a conversation progresses, as well as long-term shared knowledge to proactively build and evaluate interaction with the user over time. Intelligent Hypothesis Builder 310 may thus adaptively arrive at a set of n-best hypotheses about user intent, and the n-best hypotheses may be provided to an Adaptive Response Builder 315. In addition, Intelligent Hypothesis Builder 310 may model human-to-human interaction by calculating a degree of certainty for each of the hypotheses. That is, just as humans rely on knowledge shared by participants to examine how much and what kind of information was available, Intelligent Hypothesis Builder 310 may leverage the identified conversation type and short-term

Google Exhibit 1019
Page 00642

and long-term shared knowledge to generate a degree of certainty for each hypothesis.

[048] According to another aspect of the invention, Intelligent Hypothesis Builder 310 may generate one or more explicit hypotheses of a user's intent when an utterance contains all information (including qualifiers) needed to complete a request or task. Each hypothesis may have a corresponding degree of certainty, which may be used to determine a level of unprompted support to provide in a response. For example, a response may include a confirmation to ensure the utterance was not misunderstood. or the response may adaptively prompt a user to provide missing information.

[049] According to another aspect of the invention, Intelligent Hypothesis Builder 310 may use short-term knowledge to generate one or more implicit hypotheses of a user's intent when an utterance may be missing required qualifiers or other information needed to complete a request or task. Each hypothesis may have a corresponding degree of certainty. For instance, when a conversation begins, short-term knowledge stored in the Session Input Accumulator may be empty, and as the conversation progresses, the Session Input Accumulator may build a history of the conversation. Intelligent Hypothesis Builder 310 may use data in the Session Input Accumulator to supplement or infer additional information about a current utterance. For example, Intelligent Hypothesis Builder 310 may evaluate a degree of certainty based on a number of previous requests relevant to the current utterance. In another example, when the current utterance contains insufficient information to complete a request or task, data in the Session Input Accumulator may be used to infer missing information so that a hypothesis can be generated. In still another example, Intelligent Hypothesis Builder 310 may identify syntax and/or grammar to be used by Adaptive Response Builder 315 to formulate personalized and conversational response. In yet another example, when the current utterance contains a threshold amount of information needed to complete a request or task, data in the Session Input Accumulator may be relied upon to tune a degree of certainty.

[050] According to another aspect of the invention, Intelligent Hypothesis Builder 310 may use long-term shared knowledge to generate one or more implicit hypotheses of a user's intent when an utterance is missing qualifiers or other information needed to complete a request or task. Each hypothesis may have a corresponding degree of certainty. Using long-term knowledge may be substantially similar to using short-term shared knowledge, except that information may be unconstrained by a current session,

Google Exhibit 1019
Page 00643

and an input mechanism may include information from additional sources other than conversational sessions. For example, Intelligent Hypothesis Builder 310 may use information from long-term shared knowledge at any time, even when a new conversation is initiated, whereas short-term shared knowledge may be limited to an existing conversation (where no short-term shared knowledge would be available when a new conversation is initiated). Long-term shared knowledge may come from several sources, including user preferences or a plug-in data source (e.g., a subscription interface to a remote database), expertise of a user (e.g., based on a frequency of errors, types of tasks requested, etc., the user may be identified as a novice, intermediate, experienced, or other type of user), agent-specific information and/or language that may also apply to other agents (e.g., by decoupling information from an agent to incorporate the information into other agents), frequently used topics passed in from the Session Input Accumulator, frequently used verbs, nouns, or other parts of speech, and/or other syntax information passed in from the Session Input Accumulator, or other sources of long-term shared knowledge may be used.

[051] According to another aspect of the invention, knowledge-enabled utterances, as generated by Intelligent Hypothesis Builder 310, may include one or more explicit (supplied by a user), and one or more implicit (supplied by Intelligent Hypothesis Builder 310) contextual signifiers, qualifiers, criteria, and other information that can be used to identify and evaluate relevant tasks. At that point, Intelligent Hypothesis Builder 310 may provide an input to Adaptive Response Builder 315. The input received by Adaptive Response Builder 315 may include at least a ranked list of hypotheses, including explicit and/or implicit hypotheses, each of which may have a corresponding degree of certainty. A hypothesis may be assigned one of four degrees of certainty: (1) "sure," where contextual signifiers and qualifiers relate to one task, context and qualifiers relate to one task, and an ASR confidence level exceeds a predetermined threshold; (2) "pretty sure," where contextual signifiers and qualifiers relate to more than one task (select top-ranked task) and criteria relates to one request, and/or the ASR confidence level is below the predetermined threshold; (3) "not sure," where additional contextual signifiers or qualifiers are needed to indicate or rank a task; and (4) "no hypothesis," where little or no information can be deciphered. Each degree of certainty may further be classified as explicit or implicit, which may be used to adjust a response. The input received by Adaptive Response Builder 310 may also include a context, user syntax and/or

Google Exhibit 1019
Page 00644

grammar, context domain agent specific information and/or preferences (e.g., a travel context domain agent may know a user frequently requests information about France, which may be shared with a movie context domain agent so that responses may occasionally include French movies).

[052] According to another aspect of the invention, Adaptive Response Builder 315 may build syntactically, grammatically, and contextually sensitive "intelligent responses" that can be used with one or more agents to generate a conversational experience for a user, while also guiding the user to reply in a manner favorable for recognition. In one implementation, the intelligent responses may include a verbal or audible reply played through an output device (e.g., a speaker), and/or an action performed by a device, computer, or machine (e.g., downloading a web page, showing a list, executing an application, etc.). In one implementation, an appropriate response may not require conversational adaptation, and default replies and/or randomly selected response sets for a given task may be used.

[053] According to another aspect of the invention, Adaptive Response Builder 310 may draw on information maintained by Intelligence Hypothesis Builder 310 to generate responses that may be sensitive to context, task recognition of a current utterance, what a user already knows about a topic, what an application already knows about the topic, shared knowledge regarding user preferences and/or related topics, appropriate contextual word usage (e.g., jargon), words uttered by the user in recent utterances, conversational development and/or course correction, conversational tone, type of conversation, natural variation in wording of responses, or other information. As a result, Adaptive Response Builder 315 may generate intelligent responses that create conversational feel, adapt to information that accumulates over a duration of a conversation, maintain cross-modal awareness, and keep the conversation on course.

[054] According to another aspect of the invention, Adaptive Response Builder 315 may create a conversational feel by adapting to a user's manner of speaking, framing responses appropriately, and having natural variation and/or personality (e.g., by varying tone, pace, timing, inflection, word use, jargon, and other variables in a verbal or audible response). Adapting to a user's manner of speaking may include using contextual signifiers and grammatical rules to generate one or more sentences for use as response sets that may cooperate with the user. By taking advantage of short-term (from the Session Input

Google Exhibit 1019
Page 00645

Accumulator) and long-term (from one or more profiles) shared knowledge about how a user utters a request, the responses may be modeled using techniques used to recognize requests. Adaptive Response Builder 315 may rate possible responses statistically and/or randomize responses, which creates an opportunity to build an exchange with natural variation and conversational feel. This may be a significant advantage over existing voice user interfaces with incongruous input and output, where the input is "conversational" and the output is "computerese." The following examples may demonstrate how a response may adapt to a user's input word choices and manner of speaking:

User	**Do you know** [*mumbled words*] Seattle [*more mumbled words*]?
Voice User Interface	Did you want Seattle sports scores, weather, traffic, or news?

User	**Find me** [*mumbled words*] Seattle [*more mumbled words*]?
Voice User Interface	I *found* Seattle, did you want sports scores, weather, traffic, or news?

User	**Get me** [*mumbled words*] Seattle [*more mumbled words*]?
Voice User Interface	I've *got* Seattle, did you want me to *get* sports scores, weather, traffic, or news?

[055] According to another aspect of the invention, Adaptive Response Builder 315 may frame responses to influence a user to reply with an utterance that may be easily recognized. For example, a user may utter, "Get me the news" and a voice user interface response may be "Which of these categories? Top news stories, international news, political news, or sports news?" The response may be likely to illicit utterances from the user, such as "Top news stories" or "International news," which are more likely to result in a completed request. Thus, the responses may conform to a cooperative nature of human dialog, and a natural human tendency to "parrot" what was just heard as part of a next utterance. Moreover, knowledge of current context may enhance responses to generate more meaningful conversational responses, such as in the following exchange:

User	What's the weather like in Dallas?

Google Exhibit 1019
Page 00646

Voice User Interface	In Dallas, it's sunny and 90 degrees.
User	What theaters are showing the movie "The Fantastic Four" there?
Voice User Interface	10 theaters in Dallas are showing "The Fantastic Four." Do you want show times for a particular theater?

[056] Framing the responses may also deal with misrecognitions according to human models. For example, humans frequently remember a number of recent utterances, especially when one or more previous utterances were misrecognized or unrecognized. Another participant in the conversation may limit correction to a part of the utterance that was misrecognized or unrecognized, or over subsequent utterances and/or other interactions, clues may be provided to indicate the initial interpretation was incorrect. Thus, by storing and analyzing multiple utterances, utterances from earlier in a conversation may be corrected as the conversation progresses.

[057] According to another aspect of the invention, Adaptive Response Builder 315 may generate multi-modal, or cross-modal, responses to a user. In one implementation, responses may be aware of and control one or more devices and/or interfaces, and users may respond by using whichever input method, or combination of input methods, is most convenient. For example, a response asking the user to direct an utterance with a "Yes" or "No" in a multi-modal environment may also display alternatives visually.

[058] According to another aspect of the invention, Adaptive Response Builder 315 may correct a course of a conversation without interrupting conversational flow. Adaptive Response Builder 315 may generate intelligent responses based on the ranked list of hypotheses and corresponding degrees of certainty, which may be used to correct a course of a conversation without interrupting conversational flow. That is, even though the intelligent responses may be reasonably "sure," the intelligent responses may nonetheless sometimes be incorrect. While existing voice user interfaces tend to fail on average conversational missteps, normal human interactions may expect missteps and deal with them appropriately. Thus, responses after a misrecognition may be modeled after clarifications, rather than errors, and words may chosen in subsequent responses that move conversation forward and establish an appropriate domain to be explored with the user. For example, course correction may result in the following exchange:

Google Exhibit 1019
Page 00647

User	Can you **get** [*mumbled words here*] Mariners [*more mumbled words*]?
Voice User Interface	*I've got* the score for the Mariners game? 4-2 Mariners.
User	No, the start time for tomorrow's game.
Voice User Interface	Oh, 7:05.

[059] The above disclosure has been described in terms of specific exemplary aspects, implementations, and embodiments of the invention. However, those skilled in the art will recognize various changes and modifications that may be made without departing from the scope and spirit of the invention. Therefore, the specification and drawings are to be regarded as exemplary only, and the scope of the invention is to be determined solely by the appended claims.

Google Exhibit 1019
Page 00648

CLAIMS

What is claimed is:

1. A method for providing a voice user interface that generates cooperative conversational responses to user utterances, the responses being based on shared knowledge about a current conversation and/or one or more past conversations, the method comprising:

receiving an input, the input including at least one utterance;

updating the shared knowledge to include knowledge about the at least one utterance;

determining an intended meaning for the at least one utterance based on the updated shared knowledge; and

generating a response based on the determined intended meaning.

2. The method of claim 1, wherein updating the shared knowledge includes populating a short-term context stack with information about the at least one utterance.

3. The method of claim 2, further comprising expiring the information about the at least one utterance in the short-term context stack after a predetermined amount of time.

4. The method of claim 1, wherein updating the shared knowledge includes updating one or more long-term profiles based on information about the at least one utterance.

5. The method of claim 1, wherein determining the intended meaning includes identifying a conversation type for the at least one utterance, the conversation type based on one or more of a goal for the utterance, one or more participant roles for the utterance, and/or an allocation of information for the utterance.

6. The method of claim 1, wherein determining the intended meaning includes generating at least one hypothesis of an intended meaning for the at least one utterance, the at least one hypothesis including a degree of certainty.

7. The method of claim 6, wherein generating the at least one hypothesis includes identifying contextual signifiers and/or qualifiers in the at least one

Google Exhibit 1019
Page 00649

utterance.

8. The method of claim 7, further comprising inferring missing contextual signifiers and/or qualifiers for the at least one utterance based on the shared knowledge.

9. The method of claim 8, wherein the degree of certainty for the at least one hypothesis is based on one or more of the identified contextual signifiers and/or qualifiers, the inferred contextual signifiers and/or qualifiers, and/or contextual signifiers and/or qualifiers necessary for generating a response.

10. The method of claim 1, wherein generating the response includes adapting contextual signifiers and/or grammar in the response to reflect contextual signifiers and/or grammar in the at least one utterance.

11. The method of claim 1, wherein generating the response includes framing the response to influence a subsequent utterance.

12. The method of claim 1, further comprising:

learning that the determined intended meaning is incorrect;

further updating the shared knowledge to include knowledge that the determined intended meaning is incorrect; and

determining a second intended meaning for the at least one utterance based on the further updated shared knowledge, wherein the incorrect determined intended meaning is excluded as a possible second intended meaning.

13. The method of claim 1, wherein the input is a multi-modal input.

14. The method of claim 1, wherein the response is a multi-modal response.

15. A computer readable medium containing computer-executable instructions for providing a voice user interface that generates cooperative conversational responses to user utterances, the responses being based on shared knowledge about a current conversation and/or one or more past conversations, the computer-executable

Google Exhibit 1019
Page 00650

instructions operable when executed to:

receive an input, the input including at least one utterance;

update the shared knowledge to include knowledge about the at least one utterance;

determine an intended meaning for the at least one utterance based on the updated shared knowledge; and

generate a response based on the determined intended meaning.

16. The computer readable medium of claim 15, wherein the computer-executable instructions are operable to update the shared knowledge by populating a short-term context stack with information about the at least one utterance.

17. The computer readable medium of claim 16, wherein the computer-executable instructions are operable to expire the information about the at least one utterance in the short-term context stack after a predetermined amount of time.

18. The computer readable medium of claim 15, wherein the computer-executable instructions are operable to update the shared knowledge by updating one or more long-term profiles based on information about the at least one utterance.

19. The computer readable medium of claim 15, wherein the computer-executable instructions are operable to determine the intended meaning by identifying a conversation type for the at least one utterance, the conversation type based on one or more of a goal for the utterance, one or more participant roles for the utterance, and/or an allocation of information for the utterance.

20. The computer readable medium of claim 15, wherein the computer-executable instructions are operable to determine the intended meaning by generating at least one hypothesis of an intended meaning for the at least one utterance, the at least one hypothesis including a degree of certainty.

21. The computer-readable medium of claim 20, wherein the computer-executable instructions are operable to generate the at least one hypothesis by identifying contextual signifiers and/or qualifiers in the at least one utterance.

Google Exhibit 1019
Page 00651

22. The computer readable medium of claim 21, wherein the computer-executable instructions are operable to infer missing contextual signifiers and/or qualifiers for the at least one utterance based on the shared knowledge.

23. The computer readable medium of claim 22, wherein the degree of certainty for the at least one hypothesis is based on one or more of the identified contextual signifiers and/or qualifiers, the inferred contextual signifiers and/or qualifiers, and/or contextual signifiers and/or qualifiers necessary for generating a response.

24. The computer readable medium of claim 15, wherein the computer-executable instructions are operable to generate the response by adapting contextual signifiers and/or grammar in the response to reflect contextual signifiers and/or grammar in the at least one utterance.

25. The computer readable medium of claim 15, wherein the computer-executable instructions are operable to generate the response by framing the response to influence a subsequent utterance.

26. The computer readable medium of claim 15, wherein the computer-executable instructions are further operable when executed to:

learn that the determined intended meaning is incorrect;

further update the shared knowledge to include knowledge that the determined intended meaning is incorrect; and

determine a second intended meaning for the at least one utterance based on the further updated shared knowledge, wherein the incorrect determined intended meaning is excluded as a possible second intended meaning.

27. The computer readable medium of claim 15, wherein the input is a multi-modal input.

28. The computer readable medium of claim 15, wherein the response is a multi-modal response.

Google Exhibit 1019
Page 00652

29. A system for providing a voice user interface that generates cooperative conversational responses to user utterances, the system including at least one memory for storing shared knowledge about a current conversation and/or one or more past conversations, the system comprising:

at least one input device that receives an input, the input including at least one utterance; and

one or more processors collectively operable to:

update the shared knowledge to include knowledge about the at least one utterance;

determine an intended meaning for the at least one utterance based on the updated shared knowledge; and

generate a response based on the determined intended meaning.

30. The system of claim 29, wherein the one or more processors are operable to update the shared knowledge by populating a short-term context stack with information about the at least one utterance.

31. The system of claim 30, wherein the one or more processors are operable to expire the information about the at least one utterance in the short-term context stack after a predetermined amount of time.

32. The system of claim 29, wherein the one or more processors are operable to update the shared knowledge by updating one or more long-term profiles based on information about the at least one utterance.

33. The system of claim 29, wherein the one or more processors are operable to determine the intended meaning by identifying a conversation type for the at least one utterance, the conversation type based on one or more of a goal for the utterance, one or more participant roles for the utterance, and/or an allocation of information for the utterance.

34. The system of claim 29, wherein the one or more processors are operable to

Google Exhibit 1019
Page 00653

determine the intended meaning by generating at least one hypothesis of an intended meaning for the at least one utterance, the at least one hypothesis including a degree of certainty.

35. The system of claim 34, wherein the one or more processors are operable to generate the at least one hypothesis by identifying contextual signifiers and/or qualifiers in the at least one utterance.

36. The system of claim 35, wherein the one or more processors are operable to infer missing contextual signifiers and/or qualifiers for the at least one utterance based on the shared knowledge.

37. The system of claim 36, wherein the degree of certainty for the at least one hypothesis is based on one or more of the identified contextual signifiers and/or qualifiers, the inferred contextual signifiers and/or qualifiers, and/or contextual signifiers and/or qualifiers necessary for generating a response.

38. The system of claim 29, wherein the one or more processors are operable to generate the response by adapting contextual signifiers and/or grammar in the response to reflect contextual signifiers and/or grammar in the at least one utterance.

39. The system of claim 29, wherein the one or more processors are operable to generate the response by framing the response to influence a subsequent utterance.

40. The system of claim 29, wherein the one or more processors are further operable to:

learn that the determined intended meaning is incorrect;

further update the shared knowledge to include knowledge that the determined intended meaning is incorrect; and

determine a second intended meaning for the at least one utterance based on the further updated shared knowledge, wherein the incorrect determined intended meaning is excluded as a possible second intended meaning.

Google Exhibit 1019
Page 00654

41. The system of claim 29, wherein the input is a multi-modal input.

42. The system of claim 29, wherein the response is a multi-modal response.

Google Exhibit 1019
Page 00655

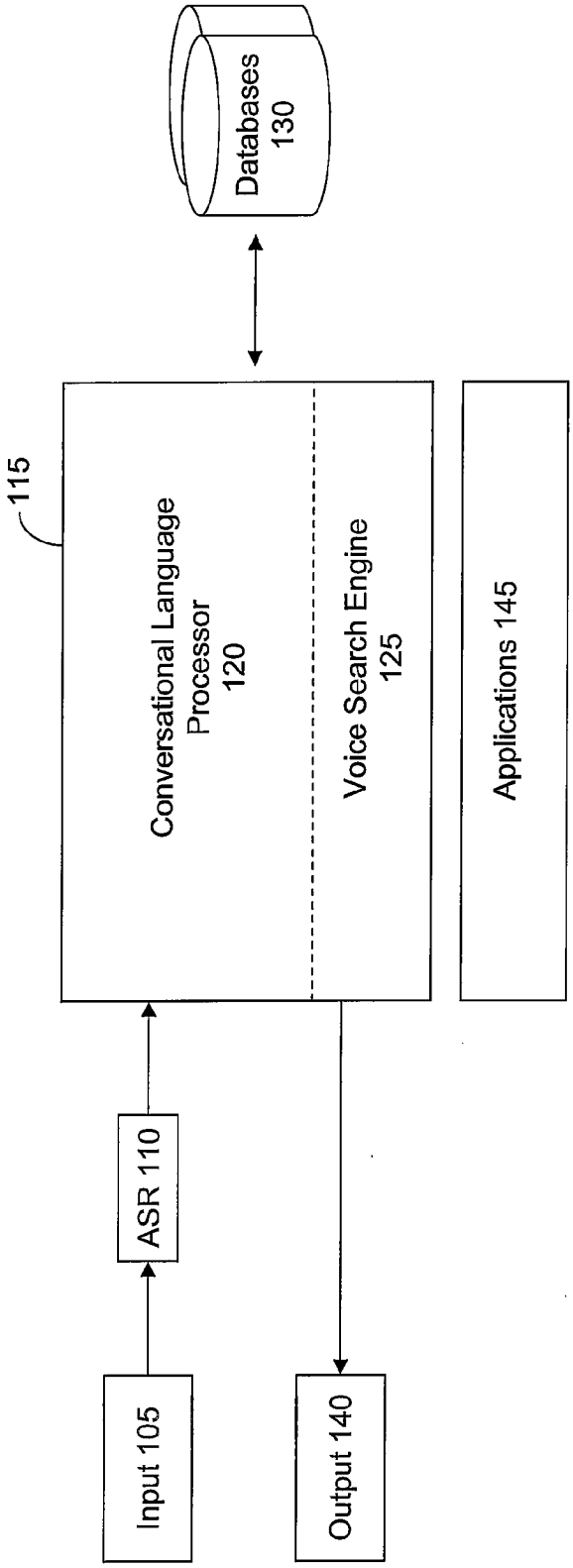

Databases
130
115
Conversational Language
Processor
120
Voice Search Engine
125
Applications 145
ASR 110
Input 105
Output 140

FIG. 1

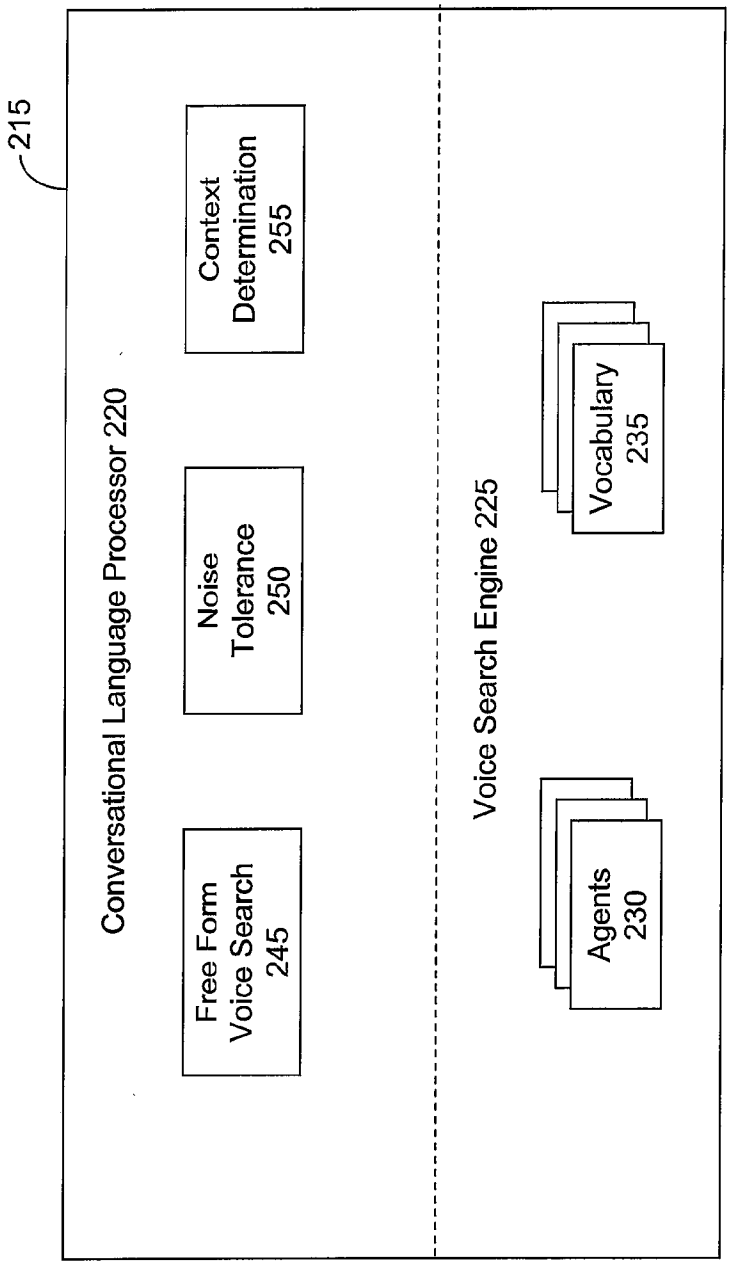

215
Conversational Language Processor 220
Free Form Voice Search 245
Noise Tolerance 250
Context Determination 255
Voice Search Engine 225
Agents 230
Vocabulary 235

FIG. 2

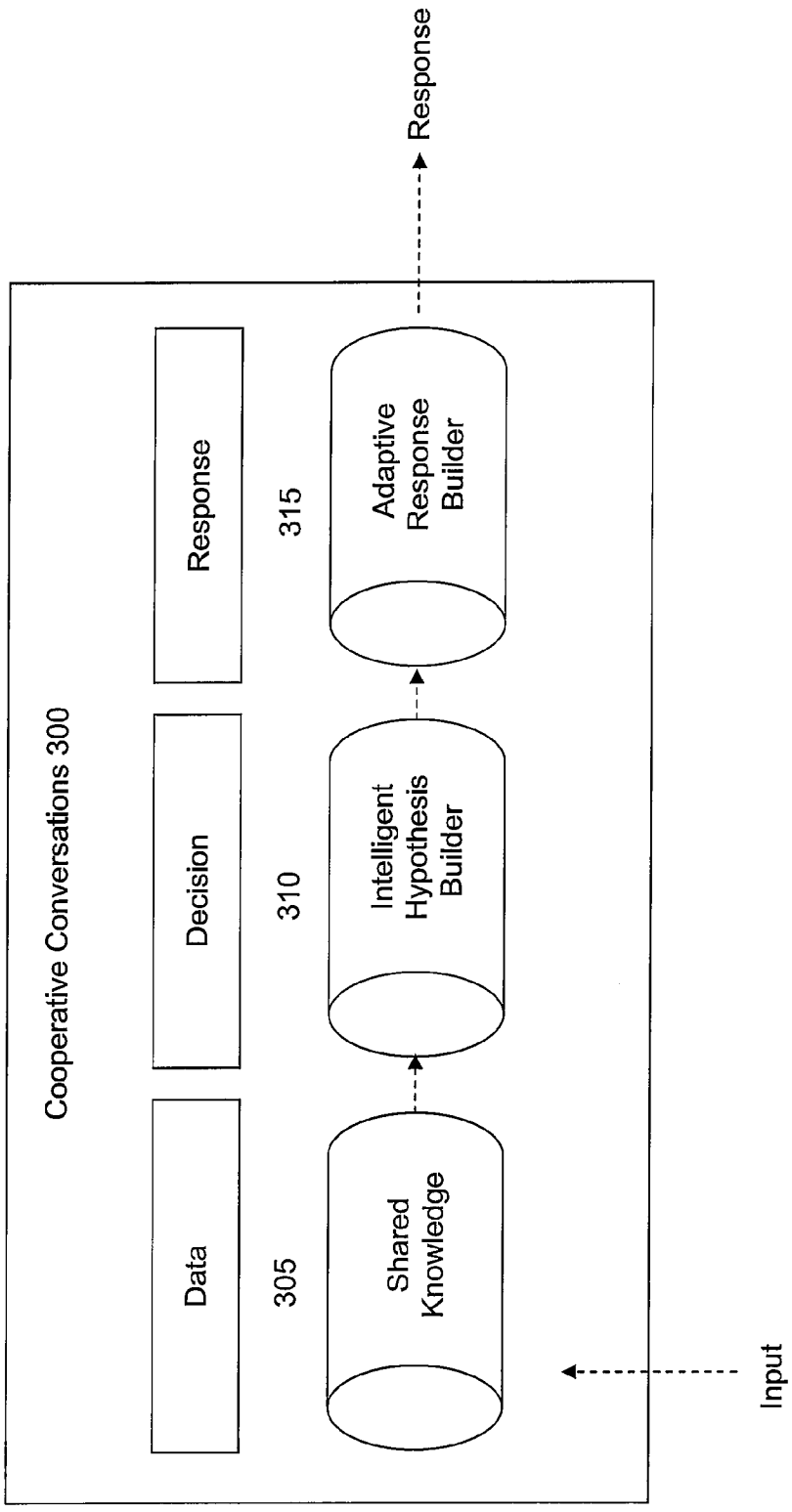

Response
Cooperative Conversations 300
Data
Decision
Response
305
310
315
Shared Knowledge
Intelligent Hypothesis Builder
Adaptive Response Builder
Input

FIG. 3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 June 2009 (18.06.2009)

PCT

(10) International Publication Number
WO 2009/075912 A1

(51) International Patent Classification:
G06F 17/27 (2006.01)

(21) International Application Number:
PCT/US2008/069508

(22) International Filing Date: 9 July 2008 (09.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/954,064 11 December 2007 (11.12.2007) US

(71) Applicant *(for all designated States except US)*: **VOICEBOX TECHNOLOGIES, INC.** [US/US]; 11980 NE 24th Street, Bellevue, Washington 98005 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KENNEWICK, Michael** [US/US]; 12196 NE 24th Street, Bellevue, Washington 98005 (US). **CHEUNG, Catherine** [US/US]; 1639 207th Avenue NE, Sammamish, Washington 98074 (US). **BALDWIN, Larry** [US/US]; 23125 SE 243rd Place, Maple Valley, Washington 98038 (US). **SALOMON, Ari** [US/US]; 4133 96th Avenue SE, Mercer Island, Washington 98040 (US). **TJALVE, Michael** [DK/US]; 14904 SE 47th Court, Bellevue, Washington 98006 (US). **GUTTIGOLI, Sheetal** [IN/US]; 11314 Ohde Circle, Kirkland, Washington 98033 (US). **ARMSTRONG, Lynn** [US/US]; 19807 183rd Place NE, Woodinville, Washington 98077 (US). **DI CRISTO, Philippe** [FR/US]; 177 107th Avenue NE, #1007, Bellevue, Washington 98004 (US). **ZIMMERMAN, Bernie** [US/US]; 135 S. 107th Street, Seattle, Washington 98168 (US). **MENAKER, Sam** [US/US]; 480 150th Place NE, Bellevue, Washington 98007 (US).

(74) Agents: **GATTO, James G.** et al.; Pillsbury Winthrop Shaw Pittman LLP, P.O. Box 10500, McLean, Virginia 22102 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: PROVIDING A NATURAL LANGUAGE VOICE USER INTERFACE IN AN INTEGRATED VOICE NAVIGATION SERVICES ENVIRONMENT

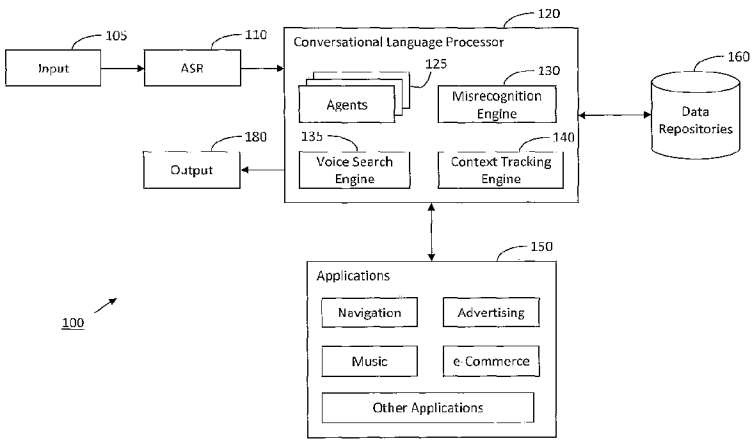


Figure 1

(57) Abstract: A conversational, natural language voice user interface may provide an integrated voice navigation services environment. The voice user interface may enable a user to make natural language requests relating to various navigation services, and further, may interact with the user in a cooperative, conversational dialogue to resolve the requests. Through dynamic awareness of context, available sources of information, domain knowledge, user behavior and preferences, and external systems and devices, among other things, the voice user interface may provide an integrated environment in which the user can speak conversationally, using natural language, to issue queries, commands, or other requests relating to the navigation services provided in the environment.

WO 2009/075912 A1

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

PROVIDING A NATURAL LANGUAGE VOICE USER INTERFACE IN AN INTEGRATED VOICE NAVIGATION SERVICES ENVIRONMENT

FIELD OF THE INVENTION

[001] The present invention relates to a natural language voice user interface that facilitates cooperative, conversational interactions in an integrated voice navigation services environment, and in particular, to a natural language voice user interface in which users can request navigation services using conversational, natural language queries or commands.

BACKGROUND OF THE INVENTION

[002] As technology advances, consumer electronics tend to play increasingly significant roles in everyday life. As a result, users tend to expect greater functionality, mobility, and convenience from their electronic devices, as exemplified by modern mobile phones, navigation devices, personal digital assistants, portable media players, and other devices often providing a wealth of functionality beyond core applications. However, the greater functionality often tends to be accompanied by significant learning curves and other barriers that prevent users from fully exploiting device capabilities (e.g., features may often be buried within difficult to navigate menus or interfaces). Moreover, although increasing demand for mobility magnifies the need for simple on-the-go device interaction mechanisms, existing systems often have complex human to machine interfaces. For example, existing human to machine interfaces tend to primarily utilize various combinations of keyboards, keypads, point and click techniques, touch screen displays, or other interface mechanisms. However, these interfaces may often be unsuitable for mobile or vehicular devices (e.g., navigation devices), as they tend to be cumbersome in environments where speed of interaction and dangers of distraction pose significant issues. As such, existing systems often fall short in providing simple and intuitive interaction mechanisms, potentially inhibiting mass-market adoption for certain technologies. As such, there is an ever-growing demand for ways to exploit technology in intuitive ways.

[003] In response to these and other problems, various existing systems have turned to voice recognition software to simplify human to machine interactions. For example, voice recognition software can enable a user to exploit applications and features of a device that may otherwise be unfamiliar, unknown, or difficult to use. However, existing voice user interfaces, when they

Google Exhibit 1019
Page 00661

actually work, still require significant learning on the part of the user. For example, existing voice user interfaces (e.g., command and control systems) often require users to memorize syntaxes, words, phrases, or other keywords or qualifiers in order to issue queries or commands. Similarly, when users may be uncertain of exactly what to request, or what a device may be capable of, existing systems cannot engage with the user in a productive, cooperative, natural language dialogue to resolve requests and advance conversations. Instead, many existing speech interfaces force users to use predetermined commands or keywords to communicate requests in ways that systems can understand. By contrast, cognitive research on human interaction demonstrates that a person asking a question or giving a command typically relies heavily on context and shared knowledge of an answering person. Similarly, the answering person also tends to rely on the context and shared knowledge to inform what may be an appropriate response. However, existing voice user interfaces do not adequately utilize context, shared knowledge, or other similar information to provide an environment in which users and devices can cooperate to satisfy mutual goals through conversational, natural language interaction.

[004] Furthermore, demand for global positional systems and other navigation-enabled devices has grown significantly in recent years. Navigation devices often tend to be used while a user may be driving, on-the-go, or in other environments where having a hands-free interface provides critical advantages. For example, a user may want to avoid being distracted by looking away from the road, yet the user may also want to interact with a navigation device, for example, to calculate a route to a destination, recalculate the route in response to traffic, find a local restaurant, gas station, or other point of interest, or perform another navigation related task. In these and other instances, efficiently processing a natural language voice-based input could enable the user to interact with the navigation device in a safer, simpler, and more effective way. However, existing systems often fall short in providing an integrated, conversational, natural language voice user interface that can provide such advantages in navigation and other mobile environments.

[005] Existing systems suffer from these and other problems.

SUMMARY OF THE INVENTION

Google Exhibit 1019
Page 00662

[006] According to various aspects of the invention, various problems associated with existing systems may be addressed by a conversational, natural language voice user interface that provides an integrated voice navigation services environment.

[007] According to various aspects of the invention, the natural language voice user interface may resolve voice requests relating to navigation (e.g., calculating routes, identifying locations, displaying maps, etc.). The navigation application can provide a user with interactive, data-driven directions to a destination or waypoint, wherein the user can specify the destination or waypoint using free-form natural language (e.g., the user can identify full or partial destinations, including a specific address, a general vicinity, a city, a name or type of a place, a name or type of a business, a name of a person, etc.). As free form voice destination inputs may be provided in many different forms, post-processing may be performed on full or partial voice destination inputs to identify a suitable destination address for calculating a route (e.g., a closest address that makes "sense"). For example, an utterance containing a full or partial destination may be analyzed to identify one or more probable destinations (e.g., an N-best list of destinations). The N-best list may be post-processed to assign weights or rankings to the probable destinations (e.g., based on a degree of certainty that a given probable destination corresponds to an intended destination). Thus, a route may be calculated from a user's current location to a most heavily weighted one of the probable destinations in the N-best list. Further, when a voice destination entry includes a partial destination, a final destination may be successively refined over one or more subsequent voice destination entries. The navigation application may also provide dynamic, data-driven directions or routing to a destination. For instance, the navigation application may access data associated with various user-specific and environmental data sources to provide personalized data-driven directions along a route, which can be recalculated or modified based on information taken from the data sources. As such, data may be obtained dynamically to identify alternate routes, recalculate routes, or otherwise provide routing services. Further, possible answers or responses to a given utterance may be filtered according to a current route.

[008] According to various aspects of the invention, the natural language voice user interface may dynamically generate and/or load recognition grammars for interpreting what was said in an utterance (e.g., content of the utterance). Information contained in the dynamic recognition grammars may be used by a navigation agent, an Automatic Speech Recognizer, a context stack,

Google Exhibit 1019
Page 00663

or various other components in the voice user interface that use grammar information. By efficiently generating, updating, loading, extending, or otherwise building dynamic grammars based on various factors, processing bottlenecks can be avoided, conflicts can be reduced, and other aspects of interpreting an utterance using a recognition grammar can be optimized. For example, a size of a generated grammar may be constrained by an amount of resources available in a system (e.g., in embedded devices or other devices having low amounts of dynamic memory, the constrained grammar size may limit a quantity of resources to be occupied). In another example, the size of the dynamic grammar can be reduced by eliminating redundant keywords, criteria, or other information available in the context stack, the shared knowledge, or other local sources. Thus, favorability of correct interpretations may be improved by reducing perplexity in the grammar (e.g., when two or more elements may likely be confused, one or more of the elements may be eliminated to reduce confusion).

[009] According to various aspects of the invention, the natural language voice user interface may generate dynamic recognition grammars using techniques of geographical chunking. A user's location can be determined at a given moment to determine one or more geographic proximities, which can be used to form an appropriate topological domain for the grammar. For example, the topological domains may reflect physical proximities (e.g., a distance from a current location), civil organization proximities (e.g., regions, states, cities, neighborhoods, subdivisions, localities, etc.), temporal proximities (e.g., amounts of travel time from the current location), directional proximities (e.g., based on directional travel vectors), or various combinations thereof. As a result, by mapping the user's geographic proximities to one or more topological domains, dynamic grammars may be pruned, extended, swapped in or out of memory, or otherwise generated and/or loaded to provide optimal recognition based on location, time, travel, or other factors (e.g., information may be swapped in and out of a grammar as a user moves from one area to another, ensuring that system resources utilize information presently relevant to a given location).

[010] According to various aspects of the invention, the natural language voice user interface may include dynamic grammars formed from one or more topological domains, which may be subdivided into a plurality of tiles, which may be further subdivided into a plurality of subtiles. Thus, information used to build the dynamic grammar can be subdivided or weighed in various ways to determine what information should be included in the grammar. Moreover,

Google Exhibit 1019
Page 00664

geographical chunks based on physical, civil organization, temporal, directional, or other proximities may be extended into other domains in which a topological taxonomy can be placed. As a result, in addition to having relevance in navigation or other location-dependent systems, the geographical chunking techniques can be applied in other contexts or domains in which geography or location may be relevant. Further, a server operably coupled to the voice user interface may analyze various forms of information to build or refine a source of grammar information. For example, when various devices communicate with the server, information may be communicated to the server may be used to update proximities, topological domains, tiles, subtiles, peer-to-peer affinities, or other grammar information.

[011] According to various aspects of the invention, the natural language voice user interface may calculate routes, provide dynamic data-driven directions to a destination, provide dynamic routing to a destination, perform post-processing of full or partial destination entries, or otherwise provide various voice navigation services. Further, destinations and/or routes may be identified using techniques of successive refinement for voice destination entries, wherein context, agent adaptation, and shared knowledge, among other things, can help a user to narrow down a final destination using voice commands, multi-modal commands, or various combinations thereof. However, it will be apparent that the successive refinement techniques can be applied to various tasks in which generalized approximations can be successively refined through voice or multi-modal commands to narrow down information sought by a user, including various other domains, contexts, applications, devices, or other components that employ the techniques described herein.

[012] According to various aspects of the invention, the natural language voice user interface may enable successive refinement of a final destination by progressively narrowing the final destination. For example, successively refining the destination may be modeled after patterns of human interaction in which a route or a destination may be narrowed down or otherwise refined over a course of interaction. For example, a user may generally approximate a destination, which may result in a route being calculated along a preferred route to the approximated destination. While en route to the approximated destination, the user and the voice user interface may cooperatively refine the final destination through one or more subsequent interactions. Thus, a user may provide a full or partial destination input using free form natural language, for example, including voice commands and/or multi-modal commands. One or more

Google Exhibit 1019
Page 00665

interpretations of a possible destination corresponding to the voice destination input may be organized in an N-best list. The list of possible destinations may be post-processed to assign weights or ranks to one or more of the entries therein, thus determining a most likely intended destination from a full or partial voice destination input. Thus, the post-processing operation may rank or weigh possible destinations according to shared knowledge about the user, domain-specific knowledge, dialogue history, or other factors. As a result, the full or partial destination input may be analyzed to identify an address to which a route can be calculated, for example, by resolving a closest address that makes "sense" relative to the input destination. Subsequent inputs may provide additional information relating to the destination, and the weighted N-best list may be iteratively refined until the final destination can be identified through successive refinement. As a result, when a suitable final destination has been identified, the route to the final destination may be completed.

[013] According to various aspects of the invention, the natural language voice user interface may include one or more advertising models for generating and/or detecting events relating to location dependent advertisements for navigation systems (e.g., as generated by a local or remote advertising engine, or via a data channel, or in other ways). For example, navigation systems typically include various mechanisms for determining a current location (e.g., a global positioning system, a radio frequency identification system, a system that determines location based on a distance to an identifiable wireless tower or access point, etc.). The location detection system may thus detect information associated with a radio frequency identifier over a data channel used by a marketer to provide advertisements. The marketer may broadcast the advertisement via the data channel, such that the navigation system triggers an event when within a suitable proximity of the RFIDs. Thus, information associated with the event may be filtered according to the current routing information or other contextual parameters to determine what action should be taken in response thereto. In other instances, advertisements may be uploaded to a server by one or more advertising partners, wherein the uploaded advertisements may be associated with metadata or other descriptive information that identifies a target audience, location-dependent information, or other criteria. In another example, a plurality of advertisements may be stored locally at the voice user interface, and an inferencing engine may determine appropriate circumstances in which an event should be generated to deliver one or more of the advertisements to a user. As a result, it will be apparent that advertising events may

Google Exhibit 1019
Page 00666

be generated in a number of ways, and may be generated and/or detected locally, remotely, by detecting RFIDs, or in other ways.

[014] According to various aspects of the invention, the natural language voice user interface may track user interactions with delivered advertisements. In this way, affinity based models may be generated, for example, to ensure that promotions or advertisements will be delivered to a likely target audience. Thus, an event relating to a given advertisement may be generated and/or detected when shared knowledge about a user's behavior, preferences, or other characteristics match one or more criteria associated with peer-to-peer affinities associated with the advertisement. In other examples, an advertising model may include mobile pay-per-use systems, peer-to-peer local guides or recommendations, or other forms of advertising. Additionally, various aspects of the advertising model, such as the local guides and recommendations, may be generated according to a mapping applied to various topological domains. For example, certain types of advertisements may be dependent on geographic or topological characteristics, and such advertisements may be associated with a topological taxonomy based on geographical chunks. As a result, various advertising events may be generated and/or detected according to physical proximities, temporal proximities, directional proximities, civil organization proximities, or various combinations thereof.

[015] According to various aspects of the invention, the natural language voice user interface may enable a user to provide requests (e.g., queries, commands, or other requests) to the navigation device using natural language. As such, the user and the navigation device may engage in a cooperative, conversational dialogue to resolve the request. For example, the voice user interface may use prior context, dialogue histories, domain knowledge, short and long-term shared knowledge relating to user behavior and preferences, noise tolerance, and cognitive models, among various other things, to provide an integrated environment in which users can speak conversationally, using natural language, to issue queries, commands, or other requests that can be understood and processed by a machine. Accordingly, the voice user interface may understand free form human utterances, freeing the user from restrictions relating to how commands, queries, or other types of requests should be formulated. Instead, the user can use a natural or casual manner of speaking to request various voice services in an integrated environment, in which various devices can be controlled in a conversational manner, using natural language. For example, the voice user interface may be aware of data and services

Google Exhibit 1019
Page 00667

associated with the navigation device, a media device, a personal computer, a personal digital assistant, a mobile phone, or various other computing devices or systems available in the environment.

[016] According to various aspects of the invention, the natural language voice user interface may include an input mechanism that receives a voice-based input, which includes at least an utterance or verbalization spoken by a user. The input mechanism may include a suitable device or combination of devices that can receive voice-based inputs (e.g., a directional microphone, an array of microphones, or other devices that encode speech). The input mechanism can be optimized to maximize gain in a direction of a user, cancel echoes, null point noise sources, perform variable rate sampling, filter out background conversations or environmental noise, or otherwise optimize fidelity of encoded speech. As such, the input mechanism may generate encoded speech generated in a manner that tolerates noise or other factors that could otherwise interfere with interpreting speech. Further, in various implementations, the input mechanism may include one or more other (non-voice) input modalities, which can be processed and/or correlated with one or more previous, current, or subsequent utterances or other voice-based inputs. As such, a user can provide other forms of input using a touch-screen interface, a stylus/tablet interface, a keypad or keyboard, or other input interfaces, for example, to clarify utterances or provide additional information about the utterances using other input modalities. For instance, the user could touch a stylus or other pointing device to a portion of a map displayed on a touch-screen interface, while also providing an utterance relating to the touched portion (e.g., "Show me restaurants around here."). In this example, the inputs can be correlated to interpret "around here" as likely referring to the touched portion of the map, as distinct from the user's current location or some other meaning.

[017] According to various aspects of the invention, the natural language voice user interface may include an Automatic Speech Recognizer that processes encoded speech to generate one or more preliminary interpretations of what was said in an utterance (e.g., content of the utterance). For example, the Automatic Speech Recognizer may generate the preliminary interpretations using phonetic dictation to recognize a stream of phonemes based on a dynamically adaptable recognition grammar. The dynamically adaptable recognition grammar may be based on dictionaries or phrases from various input domains (e.g., domains for languages, navigation, music, movies, weather, various temporal or geographic proximities, or various other domains).

Google Exhibit 1019
Page 00668

Thus, the Automatic Speech Recognizer may generate one or more interpretations of an utterance, which may be represented as a series of phonemes or syllables. The one or more interpretations can be analyzed (e.g., utilizing phonotactic rules or models of human speech) to generate an N-best list of preliminary interpretations as to what was spoken by the user. The preliminary interpretations may then be provided to a conversational language processor, which utilizes shared knowledge, contextual information, and various other sources of information to generate an intelligent hypothesis as to an actual meaning, the user's intent, or other aspects of the utterance. By formulating the hypothesis using various features and components that model everyday human-to-human conversations, the conversational language processor may generate a hypothesis as to the meaning or intent of the utterance, which can inform a process of resolving one or more requests contained in the utterance.

[018] According to various aspects of the invention, the natural language voice user interface may include, among other things, a context tracking engine that establishes meaning for a given utterance. For example, the context tracking engine can manage competitions among one or more context-specific domain agents that establish the meaning (e.g., redistributable, adaptable engines or modules that provide functionality for a given context, domain, system, or application). The domain agents may analyze preliminary interpretations of an utterance to generate a domain-specific probable interpretation. For example, one or more of the agents may include adaptable vocabularies, concepts, available tasks, or other forms of information specific to the respective domain or context. In addition, the agents can use a voice search engine to search a network for information that may not be available within the system. The probable interpretation can be assigned a weighted ranking or score, which can be used to select a "winning" one of the agents. Thus, the winning one of the agents may be designated responsible for establishing or inferring further information (e.g., based on domain or context-specific information), updating the shared knowledge, or resolving requests in the utterance, among other things. The context tracking engine may also maintain a context stack to track conversation topics, track previously invoked agents, evaluate criteria, weigh parameters, or otherwise maintain information relating to a conversational context (e.g., the context stack may be traversed in light of recent contexts, frequently used contexts, or other information included therein to determine a most likely intent of the user). By identifying a context, or correlatively, by identifying capabilities, tasks, vocabularies, or other information within the context, the

Google Exhibit 1019
Page 00669

context tracking engine can provide relevant information for establishing intent in addition to phonetic clues associated with the utterance (e.g., a word with multiple possible meanings may be disambiguated based on its meaning in a given context, previous usage in a dialogue, etc.).

[019] According to various aspects of the invention, the natural language voice user interface may utilize various forms of information to enable sharing of assumptions and expectations relating to a given utterance, conversation, or other human to machine interaction. For example, to inform decision making in the voice user interface, the voice user interface may include information sources containing short-term and long-term shared knowledge relating to user behavior, preferences, or other characteristics (e.g., short-term and long-term profiles of a specific user, global users, peer users, etc.). The short-term shared knowledge may accumulate information during a current conversation to dynamically establish awareness of a state of the voice user interface (e.g., recognition text for previous utterances, a cross-modal user interface manipulation history, a list of previously selected tasks or invoked queries, or other information). Storage of the short-term knowledge may be modeled after human interaction, such that certain information may be expired after a psychologically appropriate amount of time (e.g., to expunge stale data), whereas information with long-term significance can be added to the long-term shared knowledge (e.g., to establish persistent awareness of data likely to remain static over time). As such, the long-term shared knowledge may profile or otherwise model various characteristics, preferences, behavior, or other information relating to a user based on information accumulated over time (e.g., user-specific jargon, demographics, cognitive patterns, frequently requested tasks, favorite topics or concepts, etc.). Accordingly, the voice user interface may utilize various forms of information available via a context tracking engine, domain agents, a voice search engine, shared knowledge, internal or external databases, data associated with other devices, or other knowledge sources. As a result, a conversational type or goal associated with utterance can be identified. Based on the available information and the type or goal of the conversation, the voice user interface may attempt to resolve the utterance (e.g., by invoking an agent, which utilizes one or more applications to perform a requested task, retrieve requested information, etc.).

[020] According to various aspects of the invention, the natural language voice user interface may generate cross-modal intelligent responses sensitive to syntax, grammar, and context, which may provide a conversational feel to system-generated responses. When available, a generated

Google Exhibit 1019
Page 00670

intelligent response may present results of a resolved request (e.g., feedback relating to a task, information retrieved as a result of a query, etc.). Further, the intelligent responses may be provided across modalities, for example, using various combinations of verbal and/or non-verbal outputs (e.g., information may be presented on a display device, via an audible alert, a verbalized output, another output mechanism, or various combinations thereof). Further, verbalized components of the cross-modal response may be adapted to the user's manner of speaking (e.g., by varying tone, pace, timing, or other variables), thus creating verbal responses having natural variation and personality. The intelligent responses may also be formulated to provide an output that guides the user towards a subsequent response (e.g., a subsequent utterance) that may be more likely to be recognized. For example, when an utterance cannot be resolved because of an ambiguous context, unrecognizable words or phrases, or other factors that can result in an ambiguous or unrecognized interpretation, the intelligent response can be framed to disambiguate context or request additional information from the user to clarify a meaning of the utterance. Further, when subsequent information indicates that a given interpretation or hypothesis was incorrect, one or more previous utterances may be reinterpreted to refine context and update short-term or long-term conversational models (e.g., an utterance of "No, I meant . . ." may indicate that a previous utterance was interpreted incorrectly, in which case various previous utterances may be reinterpreted based on a correct interpretation, thus building a more accurate conversational context). As a result, the voice user interface may tolerate full or partial failure through adaptive mechanisms.

[021] According to various aspects of the invention, the natural language voice user interface may provide voice navigation services within an agent-based architecture. The architecture may include a plurality of adaptable agents (e.g., specialized software, data, content, or other information that provide functionality, behavior, services, data, and other information in a plurality of respective contextual domains), at least one of which provides navigation services (e.g., route calculation, map control, location-sensitive information, data-driven directions, etc.). As the agents process requests, the agents may autonomously react, adapt, and otherwise reconfigure to provide optimal voice services in respective domains. For example, by building context over time (e.g., by generating short and long-term profiles of a user, conversations with the user, frequent topics or preferences, etc.), the agents may automatically combine knowledge, adapt preferences, remove conflicts, or perform other adaptations to refine or otherwise optimize

Google Exhibit 1019
Page 00671

an operational framework thereof. Adaptation of the agents, which include at least a navigation agent, may occur across the plurality of agents, for example, in response to various ones of the agents resolving voice-based requests. As such, the adaptation may occur autonomously as a by-product of the agents providing voice services, generating inferences, identifying affinities (e.g., among users, peers, communities, etc.), receiving updates from external sources (e.g., using an update manager), or in other ways, as will be apparent.

[022] According to various aspects of the invention, the natural language voice user interface may include an agent-based architecture for providing voice navigation services. For example, the agent-based architecture may include one or more domain or context-specific agents, which include at least a navigation agent. The navigation agent may include, among other things, various navigation-specific content packages (e.g., dictionaries, available queries, tasks, commands, dynamic grammars, etc.), response lists (e.g., appropriate responses to commands, queries, or other requests), personality profiles (e.g., for creating a natural feel to system-generated speech), substitution lists (e.g., for substituting or transforming data into a structured form that can be understood by a target information source), or various other forms of navigation-specific information. Further, the navigation agent may be associated with pointers to local or remote data sources, parameters and operating data provided associated with other services in the architecture, or various other forms of information. For example, the data sources used by the navigation agent may include data relating to navigation, points-of-interest, traffic, events, parking, personal data, peer affinities, or various other sources of information. Further, the data sources may be populated, extended, pruned, or otherwise constructed through adaptation, analysis of various models, communication with a data service, or in other ways, as will be apparent.

[023] According to various aspects of the invention, the natural language voice user interface may include a navigation agent, which may be coupled with various sources of information, and may make use of context, communicating with various other adaptable agents and other system components to provide voice navigation services. For example, the navigation agent may use contextual information relating to a navigation domain, including tracked topics, user locations, routes traveled, previous requests, user interface states, user behaviors, preferences, demographics, or other characteristics, or various other types of contextual information. As a result, the navigation agent may have various sources of knowledge and resources available to

Google Exhibit 1019
Page 00672

resolve voice navigation requests. For example, the navigation agent may generate inferences (e.g., using an inference engine) using the available knowledge and resources to apply various rules, policies, or other inferencing techniques to generate interpretations of an utterance (e.g., phonetic fuzzy matching, inductive logic, Bayesian probability analysis, monotonic or non-monotonic reasoning, etc.). As such, the navigation agent can infer keywords or criteria not explicitly provided in the utterance, determine suitable responses to subjective or indeterminate utterances, generate events, identify peer affinities, or otherwise generate inferences for resolving navigation-related requests.

[024] According to various aspects of the invention, the natural language voice user interface may include one or more inference engines, which can generate various inferences through awareness of previous context, short-term or long-term shared knowledge, command histories, states of vehicular systems, user interface states, and various other data sources. In various implementations, one or more of the agents may be associated with a respective inference engine that can generate inferences using domain-specific knowledge, rules, policies, or other criteria. For instance, the inference engines may identify keywords or criteria missing in an utterance, infer intended meanings, autonomously suggest available tasks, or otherwise assist an associated agent in identifying queries, commands, or other requests contained in an utterance. Moreover, when information cannot be suitably resolved using information sources associated with the navigation agent, or through generating inferences, the information may be requested from one or more other agents, other devices, network information sources (e.g., via a voice search engine), or in other ways, as will be apparent. Upon identifying the information through one or more of the other sources, the requesting agent may be adapted to make the information subsequently available. Thus, various devices, applications, systems, and other components of an architecture may cooperatively share available information and services (e.g., context, dialogue histories, shared knowledge, maps, points of interest, contact lists, user or peer affinities, dynamic grammars, available applications, command histories, etc.). Accordingly, the architecture may provide an integrated voice navigation services environment in which users can speak natural language requests relating to various available contexts, domains, applications, devices, information sources, or various combinations thereof.

[025] According to various aspects of the invention, the natural language voice user interface may accept natural language voice-based inputs to control an electronic device that can provide

Google Exhibit 1019
Page 00673

navigational information, in addition to various other devices associated with an environment in which the voice user interface operates. Furthermore, various functional aspects of the voice user interface may reside at a client device, at a server, or various combinations thereof.

[026] According to various aspects of the invention, the natural language voice user interface may support multi-modal voice inputs. Thus, a given multi-modal voice input may include at least a voice component (e.g., an utterance) in addition to one or more non-voice input components (e.g., inputs provided via a keypad, a touch-screen, a stylus/tablet combination, a mouse, a keyboard, or other input modalities). As such, the non-voice input component can provide additional information or clarification relating to the utterance, adding to an amount of input information available when processing voice. For example, the user can use other input modalities to clarify a meaning of the utterance, provide additional information about the utterance, reduce a number of device interactions needed to make a given request, or otherwise provide additional in relation to a given utterance.

[027] According to various aspects of the invention, the natural language voice user interface may utilize various of cognitive models, contextual models, user-specific models, or other models to identify queries, commands, or other requests in a voice input. For example, a given input may include information relating to one or more contextual domains, one or more of which may be invoked to interpret and/or infer keywords, concepts, or other information contained in the input. Moreover, short-term and long-term shared knowledge about a user's behavior and preferences may be used in a hybrid recognition model that also considers semantic analysis and contextual inferences. For example, certain syllables, words, phrases, requests, queries, commands, or other information may be more likely to occur in a given context. Thus, the hybrid recognition model may analyze semantic patterns to resolve what was said by an utterance, and may further rely on contextual history or other information to resolve what was meant by the utterance. The hybrid recognition model may be used in conjunction with, or independently of, a peer to peer recognition model. For example, recognition models may include awareness of global usage patterns, preferences, or other characteristics of peer users, where certain keywords, concepts, queries, commands, or other aspects of a contextual framework may employed by peer users within the context.

[028] Other objects and advantages of the invention will be apparent based on the following drawings and detailed description.

Google Exhibit 1019
Page 00674

BRIEF DESCRIPTION OF THE DRAWINGS

[029] Figure 1 illustrates a block diagram of an exemplary system for providing a navigation device with a conversational, natural language voice user interface according to various aspects of the invention.

[030] Figure 2 illustrates a block diagram of exemplary agent-based architecture for providing a navigation device with a conversational, natural language voice user interface according to various aspects of the invention.

[031] Figure 3 illustrates a flow diagram of an exemplary method for dynamically generating recognition grammars for a navigation device having a conversational, natural language voice user interface according to various aspects of the invention.

[032] Figure 4 illustrates a flow diagram of an exemplary method for processing multi-modal voice-based inputs provided to a navigation device having a conversational, natural language voice user interface according to various aspects of the invention.

[033] Figure 5 illustrates a flow diagram of an exemplary method for calculating routes based on voice-based inputs provided to a navigation device having a conversational, natural language voice user interface according to various aspects of the invention.

[034] Figure 6 illustrates a flow diagram of an exemplary method for providing voice services based on a current navigation route according to various aspects of the invention.

DETAILED DESCRIPTION

[035] According to various aspects of the invention, Figure 1 illustrates a block diagram of an exemplary system 100 for providing a navigation device with a conversational, natural language voice user interface. As illustrated in Figure 1, the system 100 may enable a user to engage in a natural language, cooperative, conversational dialogue with a voice-enabled navigation device. For example, the system 100 may understand free form human utterances, freeing the user from being restricted in how commands, queries, or other types of requests should be formulated. Instead, the user can use a casual or natural manner of speaking to request various voice navigation services, among various other voice services (e.g., services relating to telematics, communications, media, messaging, external systems, marketing, information retrieval, or various other computational services). As such, the user can use system 100 to control

Google Exhibit 1019
Page 00675

navigation devices, media devices, personal computers, personal digital assistants, wireless phones, or other computing devices or systems in a conversational manner, using natural language. By making use of context, prior information, domain knowledge, short and long-term shared knowledge, noise tolerance, and cognitive models, among other things, system 100 can provide an integrated environment in which users can speak conversationally, using natural language, to issue queries, commands, or other requests that can be understood and processed by a machine. For example, in various implementations, system 100 may utilize one or more techniques as described in co-pending U.S. Patent Application 10/452,147, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed June 3, 2003, and co-pending U.S. Patent Application 10/618,633, entitled "Mobile Systems and Methods for Responding to Natural Language Speech Utterance," filed June 15, 2003, the disclosure of which are hereby incorporated by reference in their entirety.

[036] According to various aspects of the invention, system 100 may include an input mechanism 105 that receives a voice-based input, including at least one utterance or verbalization spoken by a user. The input mechanism 105 may include a device or combination of devices suitable for receiving a voice-based input (e.g., a directional microphone, an array of microphones, or other devices that encode speech). In various implementations, the input mechanism 105 can be optimized to receive the voice-based input, for example, by maximizing gain in a direction of a user, cancelling echoes, nulling point noise sources, performing variable rate sampling, filtering out background conversations or environmental noise, or performing various other techniques for maximizing fidelity of encoded speech. As such, the encoded speech generated by input mechanism 105 may be highly tolerant of noise or other factors that may potentially interfere with interpreting human speech.

[037] Further, in various implementations, the input mechanism 105 may be coupled to other input modalities, in that various forms of input other than voice can be processed and/or correlated with one or more previous, current, or subsequent voice-based inputs. For example, the input mechanism 105 may be coupled to a touch-screen interface, a stylus/tablet interface, a keypad or keyboard, or other devices or system interfaces, as will be apparent. As a result, an amount of input information potentially available for system 100 to process voice may be maximized, as the user can clarify utterances or provide additional information about utterances using other input modalities. For instance, the user could touch a stylus or other pointing device

Google Exhibit 1019
Page 00676

to a portion of a map displayed on a touch-screen interface, while also providing an utterance relating to the touched portion (e.g., "Show me restaurants around here."). In this example, system 100 may correlate the inputs to interpret "around here" as likely referring to the touched portion of the map, as distinct from the user's current location or another meaning.

[038] System 100 may also include an Automatic Speech Recognizer 110 that receives the encoded voice input and generates one or more preliminary interpretations thereof. For example, the Automatic Speech Recognizer 110 may recognize the voice-based input using phonetic dictation to recognize a stream of phonemes based on a dynamically adaptable recognition grammar. As a result, the Automatic Speech Recognizer 110 may provide out-of-vocabulary capabilities, which may be tolerant of a user misspeaking, portions of a speech signal being dropped, or other factors that could interfere with interpreting an utterance. The dynamically adaptable recognition grammar may be based on dictionaries or phrases from various input domains (e.g., domains for different languages, navigation, music, movies, weather, various temporal or geographic proximities, or various other domains). Further, performance of the Automatic Speech Recognizer 110 may be improved, for example, by pruning a search space associated with the recognition grammar (e.g., the grammar can include a linking element, such as *schwa*, to represent an unstressed, central vowel that tends to be spoken frequently, even without the user's conscious awareness). Thus, using these and other techniques, Automatic Speech Recognizer 110 may analyze an incoming encoded utterance to represent portions of the utterance as a series of phonemes or syllables, which can be further broken down into core components of an onset, a nucleus, and a coda, among other sub-categories. The series of phonemes or syllables can then be analyzed (e.g., utilizing phonotactic rules that model human speech) to identify a plurality of preliminary interpretations or best guesses (e.g., an N-best list) as to what was actually spoken by the user. It will be apparent, however, that the Automatic Speech Recognizer 110 may use various techniques to generate the preliminary interpretations of the encoded utterance, including those described, for example, in co-pending U.S. Patent Application Ser. No. 11/513,269, entitled "Dynamic Speech Sharpening," filed August 31, 2006, the disclosure of which is hereby incorporated by reference in its entirety.

[039] The plurality of preliminary interpretations generated by the Automatic Speech Recognizer 110 may be provided to a conversational language processor 120, which utilizes shared knowledge to generate an intelligent hypothesis of an actual meaning, a user's intent, or

Google Exhibit 1019
Page 00677

other aspect of the voice-based input. The conversational language processor 120 may formulate the hypothesis using various features and components that collectively operate to model everyday human-to-human conversations.

[040] For example, the conversational language processor 120 may include a context tracking engine 140 that establishes meaning for a given utterance by, among other things, managing a competition among one or more context-specific domain agents 125 (e.g., redistributable, adaptable packages or modules that provide functionality for a given context, domain, system, or application). For example, the preliminary interpretations may be ranked by the Automatic Speech Recognizer 110, and the agents 125 may further analyze the preliminary interpretations to generate a weighted ranking or score, which can be used to select a "winning" one of the agents 125. The winning one of the agents 125 may then be responsible for establishing or inferring further information, updating the shared knowledge, or performing other tasks to aid in generating the intelligent hypothesis. Moreover, the context tracking engine 140 may use a context stack to track conversation topics, evaluate criteria, weigh parameters, or otherwise maintain contextual information for generating the hypothesis based on a conversational context (e.g., a context stack may be traversed in light of recent contexts, frequently used contexts, or other information included therein to determine a most likely intent of the user). By identifying a context, or correlatively, by identifying capabilities, tasks, vocabularies, or other information within the context, the context tracking engine 140 can provide highly relevant information for establishing intent, apart from meager phonetic clues.

[041] Furthermore, the conversational language processor 120 may utilize various other forms of knowledge to inform the generation of the intelligent hypothesis. For example, various agents 125 may adaptively include domain-specific or context-specific vocabularies, concepts, available tasks, or other forms of information relevant to the respective domain or context. In addition, the various components associated with the conversational language processor 120 can invoke a voice search engine 135 (e.g., an engine that searches a network for information) to resolve information that may not be internally available (e.g., when an external knowledge source may be useful in clarifying an intent behind a particular word, command, query subject, or other aspect of an utterance).

[042] System 100 may also enable the user and the system 100 to share assumptions and expectations relating to a given utterance, conversation, or other interaction. For example, the

Google Exhibit 1019
Page 00678

conversational language processor 120 may be coupled to one or more data repositories 160 that store short-term and long-term shared knowledge that inform decision making in the conversational language processor 120. The short-term shared knowledge may accumulate information during a current conversation (e.g., recognition text for previous utterances, a cross-modal user interface manipulation history, a list of previously selected tasks or invoked queries, or other information), thus dynamically establishing awareness of a cross-modal state of the voice user interface. Storage of the short-term knowledge may be modeled after human interaction, thus certain data may be expired after a psychologically appropriate amount of time (e.g., to expunge stale data), and information with long-term significance can be added to long-term shared knowledge (e.g., a new address of a long-term contact of the user). As such, the long-term shared knowledge may profile or otherwise model environmental, cognitive, historical, demographic, or other aspects of a user based on information accumulated over time (e.g., user-specific jargon, frequently requested tasks, favorite topics or concepts, etc.).

[043] Accordingly, the conversational language processor 120 includes various features that can be used to generate intelligent hypotheses as to a user's intent in a given utterance. For example, the hypothesis may be based on information provided via the context tracking engine 140, the agents 125, the voice search engine 135, the shared knowledge, or other knowledge sources. As a result, the conversational language processor 120 may attempt to identify a conversational type or goal of the utterance (e.g., a query for retrieving a discrete piece of information, a didactic interaction for clarifying information provided by the voice user interface, or an exploratory interaction in which conversational goals may be improvised as the conversation progresses, etc.). Based on the information available and the type or goal of the conversation, the generated hypothesis can be assigned a degree of certainty, which can inform how conversational language processor 120 resolves the utterance. For example, when the degree of certainty indicates that sufficient information has been identified, the conversational language processor may invoke one or more of the agents 125, which in turn can utilize one or more applications 150 to perform a requested task (e.g., a task relating to a navigation application, an advertising application, a music application, an electronic commerce application, or other suitable applications or tasks). In another example, one or more of the agents 125 may query the data repositories 160 or the voice search engine 135 to retrieve requested information, or otherwise take action to resolve a request for information contained in the utterance.

Google Exhibit 1019
Page 00679

[044] Additionally, the conversational language processor 120 may generate cross-modal intelligent responses, which may be syntactically, grammatically, and contextually sensitive, thus providing a conversational feel to system-generated responses. When available, the intelligent response may present results of a performed task or an executed query to the user, and the response may be provided across modalities. For example, verbal and/or non-verbal outputs 180 may be used separately or in concert (e.g., by presenting information using a display device, an audible alert, a verbalized output, another output mechanism, or various combinations thereof). Further, a verbalized component of the cross-modal output 180 may be adapted to the user's manner of speaking (e.g., by varying tone, pace, timing, or other variables), thus creating verbal responses having natural variation and personality.

[045] The intelligent responses may also be formulated to provide an output 180 that guides the user towards a subsequent response likely to be favorable for recognition. For example, when the degree of certainty reflects an ambiguous context (e.g., when the competition results in a deadlock among the agents 125), an adaptive misrecognition engine 130 may identify the utterance as ambiguous or unrecognized, and the intelligent response can be framed to disambiguate context, or request a subsequent response from the user to clarify the meaning of the utterance, for example. Further, when the conversational language processor 120 or the misrecognition engine 130 determines that a given interpretation or hypothesis was incorrect, one or more previous utterances may be reinterpreted to refine context and build more accurate short-term or long-term conversational models (e.g., an utterance of "No, I meant . . ." may indicate that a previous utterance was interpreted incorrectly, in which case various previous utterances may be reinterpreted based on a correct interpretation, thus building a more accurate conversational context). As a result, the conversational language processor 120 may recover from full or partial failure. Additional techniques for adaptively responding to misrecognition or ambiguity may be included, such as those described in co-pending U.S. Patent Application Ser. No. 11/200,164, entitled "System and Method of Supporting Adaptive Misrecognition in Conversational Speech," filed August 10, 2005, the disclosure of which is hereby incorporated by reference in its entirety.

[046] Additional information relating to the various techniques described herein, as well as other techniques that system 100 uses to provide a conversational, natural language interaction may be provided, for example, in co-pending U.S. Patent Application Ser. No. 11/197,504,

Google Exhibit 1019
Page 00680

entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed August 5, 2005, co-pending U.S. Patent Application Ser. No. 11/212,693, entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," filed August 29, 2005, and co-pending U.S. Patent Application Ser. No. 11/580,926, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed October 16, 2006, the disclosures of which are hereby incorporated by reference in their entirety.

[047] As such, system 100 may provide an environment where conversational, natural language interactions can occur between a user and the system 100. Further, as will be described in greater below, system 100 can be implemented to provide the conversational, natural language interactions, for example, as a voice user interface to one or more electronic devices or applications, including a navigation device. As will also be apparent from the further descriptions provided herein, the system 100 has capabilities for cross-device and cross-modal awareness, such that system 100 provides an environment in which the user can engage in a cooperative, natural language dialogue to request voice services relating to many different devices or applications.

[048] According to various aspects of the invention, Figure 2 illustrates a block diagram of exemplary agent-based architecture 200, which can provide a conversational, natural language voice user interface to a navigation device. The architecture 200 may include a plurality of agents 225a-n, which include specialized software, data, content, or other information that provide functionality, behavior, services, data, and other information in a plurality of respective contextual domains. The architecture 200 may be an integrated and dynamically adaptable environment, in which the agents 225 autonomously react, adapt, and reconfigure to provide optimal service in the environment. For example, by building context over time (e.g., by generating short and long-term profiles of a user, conversations with the user, frequent topics or preferences, etc.), the navigation agent 225a may automatically combine knowledge, adapt preferences, remove conflicts, or perform other adaptations to refine or otherwise optimize an operational framework thereof.

[049] For example, the agents 225, which include at least a navigation agent 225a, may adapt through ongoing use of short-term shared knowledge 270 and long-term shared knowledge 275 about a user's behavior, preferences, or other characteristics. Further, agent adaptation may occur across the plurality of agents 225, for example, in response to various ones of the agents

Google Exhibit 1019
Page 00681

225 resolving voice-based requests (e.g., by invoking one or more applications 250, querying one or more data sources 260, cooperating with one or more other agents 225, or learning through a misrecognition engine 230, a voice search engine 235, a context tracking engine 24, etc.). In various implementations, adaptation may occur autonomously as a by-product of the agents 225 providing voice services, generating inferences, identifying affinities (e.g., among users, peers, communities, etc.), receiving updates from external sources (e.g., an update manager can update content, behavior, data, commands, domain knowledge, keywords, concepts, dictionaries, or other information for one or more of the agents 225), or in other ways, as will be apparent.

[050] As illustrated in Figure 2, the architecture 200 provides the conversational, natural language voice user interface to the navigation device by way of at least a navigation agent 225a. The navigation agent 225a can include, among other things, various navigation-specific content packages (e.g., dictionaries, available queries, tasks, commands, dynamic grammars, etc.), response lists (e.g., appropriate responses to commands, queries, or other requests), personality profiles (e.g., for creating a natural feel to system-generated speech), substitution lists (e.g., for substituting or transforming data into a structured form that can be understood by a target information source), or various other forms of navigation-specific information. Further, the navigation agent 225a may be associated with parameters and operating data provided by other services in the architecture 200 (e.g., available devices, applications 250, agents 225, etc.), pointers to local or remote data sources (e.g., short-term shared knowledge 270, long-term shared knowledge 275, data sources 260, voice search engine 235, etc.), among various other forms of information.

[051] The data sources 260 used by the navigation agent 225a may include, among other things, data relating to navigation (e.g., maps, addresses, street names, directories, alternate routes, etc.), points-of-interest (e.g., restaurants, hotels, museums, tourist attractions, gas stations, etc.), traffic (e.g., dynamic road conditions, traffic, detours, or other traffic-related information), events (e.g., sporting events, concerts, protests, etc.), parking (e.g., parking garages, restricted areas or times, street lots, street parking, etc.), personal data (e.g., telephone numbers and addresses associated with a user's contact lists), peer affinities (e.g., recommendations based on similarities among users having similar preferences, demographic profiles, or other characteristics). The data

Google Exhibit 1019
Page 00682

sources 260 may be populated in various ways, such as being based on one or models, received via a data service, extended or refined through adaptation, or in other ways, as will be apparent.

[052] Further, contextual information relating to a navigation domain may be maintained via the context tracking engine 240, short-term shared knowledge 270, and long-term shared knowledge 275, among other things. For example, the contextual information may relate to tracked topics, user location, routes traveled, previous requests, user interface state, user behavior, preferences, demographics, or other characteristics, among other types of contextual information. For example, context may be built and utilized using techniques described in greater detail above with reference to Figure 1, or in the above-referenced co-pending U.S. Patent Applications, or various combinations thereof. Moreover, when navigation agent 225a cannot find information locally within architecture 200, and the information cannot be inferred, the information may be requested from one or more other agents 225b-n (e.g., to request domain-specific information), other devices, the voice search engine 235 (e.g., to search a network), external sources (e.g., the update manager may be invoked to download available updates, which may have the information), or other sources, as will be apparent. If the information can be located through one or more of the other sources, the navigation agent 225a may be adapted, extended, or otherwise updated to make the information available subsequently.

[053] Accordingly, the navigation agent 225a may be coupled with various sources of information, and may make use of context, while communicating with various other adaptable agents 225b-n and other system components to provide voice navigation services. Thus, context, shared knowledge, maps, points of interest, contact lists, user or peer affinities, dynamic grammars, available applications, or various other aspects of the navigation architecture 200 can be shared or otherwise made available to various components, devices, or aspects of the architecture 200. As a result, an inference engine 255 included with the navigation agent 225a may various knowledge sources and other resources available to provide an integrated voice navigation services environment. For example, the inference engine 255 may generate inferences from the available knowledge and resources by applying one or more rules, policies, or other inferencing mechanisms to generate probable interpretations of utterances in the environment (e.g., phonetic fuzzy matching, inductive logic, Bayesian probability analysis, monotonic or non-monotonic reasoning, etc.).

Google Exhibit 1019
Page 00683

[054] For instance, as described in greater detail above, the navigation agent 225a and one or more other agents 225b-n may compete in analyzing one or more preliminary interpretations of an utterance to generate one or more respective probable interpretations of the utterance. The navigation agent 225a may then be invoked to process the probable interpretation (e.g., by resolving information contained in the utterance) upon being identified as a "winning" one of the competing agents. For example, which of the competing agents wins may depend on context or other information contained in the utterance, whereby the navigation agent 225a may win and recalculate a route in response to an utterance of "This traffic is terrible, maybe back roads," while a music agent may win and change a radio channel in response to an utterance of "Traffic is terrible, maybe some Oldies," while a movie agent may win and search for movie show times in response to an utterance of "Traffic was terrible, maybe a comedy." Thus, when the navigation agent 225a generates a probable interpretation that results in being a winning one of the agents 225, the navigation agent 225a may manage processed for searching, inferring, or otherwise identifying information in light of context to provide voice navigation services using one or more applications 250 (e.g., navigation, grammar generation, location determination, response generation, advertising, device control, or other applications available within architecture 200).

[055] For example, the navigation agent 225a may utilize the inference engine 255 to infer keywords or criteria not explicitly provided in the utterance, determine suitable responses to subjective or indeterminate utterances (e.g., selecting a most likely answer to a query, asking a user for more information, etc.), generate events, identify peer affinities, or otherwise generate inferences for resolving navigation-related requests. The inference engine 255 may generate such inferences through awareness of previous context (e.g., through information provided by the context tracking engine 240), short-term or long-term shared knowledge 270, command histories, states of vehicular systems, user interfaces, or other devices, data sources 260, or other available information. For example, in an exemplary illustration, previous context, shared knowledge, and other information sources may be utilized in conjunction with the inference engine 255 and/or various other components of architecture 200 to enable a human to machine interaction that may occur as follows:

User:	"Show me restaurants around here."

Google Exhibit 1019
Page 00684

Voice User Interface:	"All I can find is fast food, which you don't like."
User:	"Okay, how about closer to the hotel?"
Voice User Interface:	"Found an Indian restaurant with good reviews. Want me to call them?"
User:	"Yeah go ahead, that sounds good."

[056] In the above illustrated example, the first utterance establishes a context of "restaurants," and the voice user interface responds to the utterance in light of a shared knowledge inference indicating that the user does not like fast food. Furthermore, while the second utterance does not explicitly reference a restaurant, the context established in the first utterance can be used to infer the unreferenced information, forming an unambiguous request out of what would otherwise be ambiguous. Further still, long-term shared knowledge 275 about the user (e.g., the user may frequently go to Indian restaurants) and short-term knowledge 270 about the user (e.g., a hotel associated with a current route) can narrow a search space for resolving the second utterance. Moreover, the voice user interface may invoke the response generation application to advance the conversation towards a resolution (e.g., by prompting the user to call the restaurant). Because the integrated environment has cross-device awareness, the user's subsequent response can result in the device control application formulating an appropriate command structure to dial the user's mobile phone, which may or may not be distinct from the navigation device. Accordingly, as illustrated in this example, robust awareness of context, shared knowledge, available applications, and available systems or devices, among other things, has enabled the user to use natural language to locate a restaurant along a current route matching the user's preferences, while also enabling the user to dial a phone to call the restaurant, without the user having to manually interact with devices or otherwise interact with cumbersome human to machine interfaces.

[057] It will be understood, however, that the above illustrated dialogue provides but one example of a cooperative conversation that can occur between a user and a system that includes the above-described architecture 200. Because the architecture 200 supports free form interactions, and can tolerate serendipitous variations in human speech, vocabulary usage, or other factors, it will be understood that no two conversational interactions between the user and the system will necessarily be alike. As such, the user can seamlessly switch among contexts,

Google Exhibit 1019
Page 00685

allowing subsequent utterances to occur in various available domains (e.g., a navigation-related utterance of "Take me to the Seahawks game" may be followed up with subsequent cross-domain utterances, such as "Are there tickets left?" or "When does it start?"). Thus, although various descriptions illustrate exemplary ways in which the agent-based architecture 200 can be used for voice navigation services, it will be apparent that, in a given utterance, the user can provide natural language voice-based requests relating to various available contexts, domains, applications, devices, information sources, or various combinations thereof.

[058] The navigation application available within architecture 200 may resolve natural language, voice-based requests relating to navigation (e.g., calculating routes, identifying locations, displaying maps, etc.). The navigation application can provide a user with interactive, data-driven directions to a destination or waypoint, wherein the user can specify the destination or waypoint using free-form natural language (e.g., the user can identify full or partial destinations, including a specific address, a general vicinity, a city, a name or type of a place, a name or type of a business, a name of a person, etc.). As a result, in various implementations, the navigation application may perform post-processing on the full or partial destinations to identify a suitable address for calculating a route (e.g., a closest address that makes "sense"). For example, an utterance containing a full or partial destination may be analyzed to identify one or more probable destinations (e.g., an N-best list of destinations), which may be subject to post-processing that weighs the identified probable destinations based on various factors. Thus, the probable destination most likely to provide a suitable (preliminary or complete) route may be identified from the N-best list.

[059] For example, in an illustrative conversation between a user and the voice user interface, the user may utter a voice-based input of "Take me to Seattle." Based on a current location of the user (e.g., as determined using one or more navigation sensors, radio frequency identifiers, local or remote map databases, etc.), the navigation application may select an address in Seattle that provides a route from the current location to Seattle (e.g., a central point in Seattle may be selected for long distance travel, whereas an address in northern Seattle may be selected for a current location being close to north Seattle). Further, when the requested destination does not specify a final destination (e.g., as in the above example), the user may successively refine the final destination in subsequent requests, and the post-processing may continue to select additional addresses based on the current location, a current route, or other factors. For instance,

Google Exhibit 1019
Page 00686

continuing the above-provided example, a subsequent input of "Take me to Pike Street" may result in the post-processing analyzing the current route to determine an appropriate address on Pike Street, and possibly recalculating the route, as necessary. Thus, the current route may have the user driving north-bound on Interstate-5, such that an address may be selected on Pike Street on a same side as north-bound lanes of Interstate-5 (e.g., preserving current routing information). The final destination may continue to be successively refined, in combination with the post-processing. As a result, a user knowing only that a destination falls along a 500-block of Pike Street can specify, "501 Pike Street," and even if such an address does not exist, the post-processing may find a closest address that makes sense, thus routing the user to the 500-block of Pike Street.

[060] In addition, the navigation application may provide dynamic, data-driven directions to the destination or dynamically provide routing information, among other things. Furthermore, the response generation application may work together with the navigation application and/or one or more other applications to generate system responses personalized for the user. For instance, the navigation application may access data associated with various user-specific and environmental data sources to provide personalized data-driven directions along a route, which can be recalculated or modified based on information taken from the data sources. Thus, in one example, the response generation application may generate personalized data-driven directions as follows:

Voice User Interface:	"Go to Bob's House."
Voice User Interface:	"Take a left on 14th Street."
Voice User Interface:	"Make a right here to get on the highway, and prepare for heavy traffic."

[061] In the above-illustrated example, the response has been personalized by framing directions in terms of one of the user's contacts (i.e., Bob), while an available source of traffic data has been used to provide data-driven directions. In another example, data that may affect routing can be identified and used to recalculate a route or identify alternate routes (e.g., weather data may be used to avoid snowy roads, or event data may be used to avoid a route along which a protest may be occurring). As such, data may be obtained dynamically to identify alternate

Google Exhibit 1019
Page 00687

routes, recalculate routes, or otherwise provide optimal routing service. In various implementations, the source of data may be based on persistent models, dynamically received data (e.g., a given highway may regularly have high volume during rush hour or sporting events), or other information. Routes may be recalculated automatically (e.g., user preferences may dictate that heavy traffic be avoided, or sporting event traffic may not be avoided when the user has tickets to the event), or the user can request recalculation through additional utterances (e.g., "I don't want to sit in traffic, can we go a different way?"). Further, possible answers or responses to a given utterance may be filtered according to a current route (e.g., possible results provided for an utterance of "Find me a gas station" may be filtered according to those gas stations within a given proximity of the route). Moreover, continuing the above-provided example, the user may later return to a point of origin, and personalized responses may include return directions, such as "Go back the way you came on 14th Street." Thus, the personalized responses may also create a natural feel to the directions, and may be based on context that builds over time.

[062] The navigation architecture 200 may be used in conjunction with an electronic device capable of accepting voice-based inputs and providing navigation-related information. For example, architecture 200 may be implemented within a handheld device (e.g., personal navigation device, mobile phone, etc.), a telematics devices (e.g., a vehicular navigation system), or various other suitable devices. Furthermore, various functional aspects of the architecture 200 may reside at a client device, at a server, or various combinations thereof. For example, a voice-based input may be received by the client device (e.g., a personal navigation device), which may communicate with the server to retrieve information, perform a task, process the input, or for other reasons. The server may subsequently return information to the client device in response to the request. In various implementations, however, the voice-based input may be processed within the system (e.g., processing delays may be reduced by minimizing communication with the server). Other suitable arrangements may be used, as will be apparent, and specific arrangements as to how information can be processed, retrieved, or returned to the user, among other things, may be highly dependent on context or specific implementations (e.g., what information can be obtained locally, remotely, or otherwise, what services a user subscribes to, etc.).

Google Exhibit 1019
Page 00688

[063] According to various aspects of the invention, Figure 3 illustrates a flow diagram of an exemplary method 300 for dynamically generating and/or loading recognition grammars in a conversational, natural language voice user interface. For example, a grammar generator application (e.g., as illustrated in Figure 2) may dynamically generate and/or load recognition grammars for use by a navigation agent, an Automatic Speech Recognizer, a context stack, or other components in the voice user interface that incorporate a recognition grammar. While grammar generating method 300 may be used by various components of the natural language voice user interface, the method 300 may be particularly useful in reducing processing overhead and reducing perplexity or confusion. For example, by efficiently generating, updating, loading, extending, or otherwise building dynamic grammars based on various factors, as will described in greater detail herein, bottlenecks can be avoided, conflicts can be reduced, or other aspects of interpreting utterances using a grammar can be optimized.

[064] For example, in various implementations, grammar generation may be keyed to an amount of available resources in a system that includes the voice user interface. Thus, at an operation 310, the amount of system available resources may be calculated, and a subsequent operation 320 may constrain a size of the dynamic grammar based on the calculated resources. For instance, in embedded devices or other devices having low amounts of dynamic memory, the constrained grammar size may ensure that only as many resources as necessary will be occupied. Thus, throughput in the low-memory device will not be degraded by a large-list grammar over-occupying processing capacity. In another example, the size of the dynamic grammar can be reduced by eliminating redundant keywords, criteria, or other information available in the context stack, the shared knowledge, or other local sources. However, it will be apparent that various other techniques may be used to constrain grammar size, and that effectively considering available system resources can provide processing improvements in various environments.

[065] In a navigational context, in which a user's location can change from one moment to another, grammar generation may be further optimized using techniques of geographical chunking. For instance, a user's location can be determined at a given moment (e.g., via navigation sensors, global positioning systems, etc.), and an operation 330 can use the determined location to determine one or more geographic proximities associated with the location. Subsequently, at an operation 340, the determined proximities may be used to identify one or more topological domains appropriate for the determined proximities.

Google Exhibit 1019
Page 00689

[066] As such, the proximities determined in operation 330 may be weighed or analyzed in various ways to determine how to utilize the topological domains identified in operation 340 in generating the dynamic grammar. For example, the topological domains may reflect physical proximities (e.g., distances from a current location, such as everything within a five-mile radius), civil organization proximities (e.g., regions, states, cities, neighborhoods, subdivisions, localities, etc.), temporal proximities (e.g., amounts of travel time from the current location, where road classes or other information may be used to estimate fastest/safest routes, such as highways and interstates being faster than city or surface roads), directional proximities (e.g., based on directional travel vectors, where a point ten miles ahead may be considered "closer" than a point one mile behind based on a direction of travel, an amount of time to turn around, etc.), or various combinations thereof. As a result, by mapping the user's geographic proximities to one or more topological domains, dynamic grammars may be pruned, extended, swapped in or out of memory, or otherwise generated and/or loaded to provide optimal recognition based on location, time, travel, or other factors (e.g., information may be swapped in and out of a grammar as a user drives from one area to another, optimizing the information along the drive to ensure that the system resources utilize information presently relevant for a given location).

[067] Furthermore, the topological domains may be subdivided into a plurality of tiles, which may in turn be subdivided into a plurality of subtiles. For example, a civil organization topological domain may include a tile representing grammar information for a state, and the tile may include one or more subtiles for counties within the state. Further, the county subtiles may include subtiles for various cities, neighborhoods, or other organizational boundaries within respective counties. Thus, the geographic proximities and topological domains used to build the dynamic grammar may be based on a variety of factors, and can be subdivided or weighed in various ways to determine what information to include in the grammar. Moreover, geographical chunks based on physical, civil organization, temporal, directional, or other proximities may be extended into various other domains in which a topological taxonomy can be placed. As a result, the geographical chunking techniques may have particular relevance in navigation or other location-dependent voice recognition systems, yet the techniques can be suitably applied in various contexts or domains in which geography or location may have relevance.

[068] As various implementations generate the recognition grammars dynamically, for example, to preserve system resources, information available to extend, update, or otherwise

Google Exhibit 1019
Page 00690

update the grammar may be stored remotely (e.g., on a server). Thus, various implementations may include a system having a network connection, a data service, or another communication mechanism for establishing a link to the remote source. Based on the proximities and/or the topological domains identified for a given location, context, utterance, or otherwise, one or more grammar tiles and/or subtiles may be downloaded from the remote source at an operation 350. Further, the remote source may be adapted to store, analyze, or otherwise process certain information to refine the grammar information stored therein. For example, when a given device communicates with the remote source at operation 350, the remote source may receive identifying information relating to a user of the device (e.g., user preferences, user characteristics, request characteristics, etc.), requested information, or other information. Based on the received information, the remote source may dynamically update various tiles and/or subtiles, build affinities (e.g., among users, peers, communities, etc.), or perform other actions to refine a process by which relevant information can be identified (e.g., the remote source may determine that various different users, who share certain common demographics, may be interested in particular events, locations, etc.).

[069] Thus, the grammar information downloaded at operation 350 may be based on a variety of factors, including information in a current voice-based input, current geographical proximities of the user, topological domains or geographical chunks related to the current geographical proximities, or various kinds of affinities, among other things. Based on these and other factors, appropriate information may be retrieved from various local or remote data sources (e.g., a local repository may contain a great deal of grammar information, a portion of which may be used by the dynamic grammar to interpret an utterance). The retrieved information may then be used to generate the dynamic grammars at an operation 360. As a result, the dynamic grammars may be adapted to include information favorable for correctly interpreting a given utterance. Furthermore, at an operation 370, the favorability of correct interpretations may further be improved by post-processing the generated grammars using one or more optimization techniques, such as reducing perplexity in the grammar. For example, when the grammar generated at operation 360 includes two or more elements likely to be confused (e.g., two or more streets having a same name), the optimization operation 370 may eliminate one or more of the elements to reduce confusion (e.g., selecting the element closest to a user's current location, or other criteria). In another example, when an interpretation turns out to be incorrect (e.g., a

Google Exhibit 1019
Page 00691

subsequent utterance may be "No, 14th Street Northwest, not Northeast"), the dynamic grammar may be pruned to eliminate information associated with the incorrect interpretation (e.g., information associated with Northeast may be removed). As a result, the optimization operation 370 can reduce overhead, assure that mistakes will not be repeated, and/or improve accuracy of interpretations, among other things.

[070] Although various exemplary operations have been described for dynamically extending, pruning, loading, or otherwise generating recognition grammars, it will be apparent that in operation, the natural language voice user interface may use various forms of information or techniques to determine an optimal grammar. For instance, in addition to being based on available system resources, likelihood of perplexity, geographical proximities, or user affinities, the grammar may be generated in response to conversational contexts, shared knowledge about a user, and/or adaptable agent behavior, among other factors. Accordingly, the dynamic grammar can be structured to recognize queries, commands, or other information in an utterance based on information most likely to be relevant for the utterance, with minimal ambiguity. Further, when the utterance cannot be resolved using the information currently in the grammar, various information sources or techniques may be available to update the grammar in a way that will be favorable for recognition. Thus, the above-described aspects and features of the grammar generation method 300 should be regarded as exemplary only, as many variations will be possible without departing from the inventive concepts described.

[071] According to various aspects of the invention, Figure 4 illustrates a flow diagram of an exemplary method 400 for processing multi-modal voice inputs using a conversational, natural language voice user interface. As described in connection with Figure 1 above, method 400 may be performed in a voice user interface having a mechanism for receiving voice inputs (e.g., a directional microphone, an array of microphones, or other devices that encode speech), wherein the voice inputs include at least a voice component (e.g., an utterance). The utterance, which may be received in an operation 410, may be provided using natural language or other free form manner of expression.

[072] Furthermore, to enable processing of multi-modal voice inputs, the voice user interface may have one or more input mechanisms for receiving non-voice input components (e.g., inputs provided via a keypad, a touch-screen, a stylus/tablet combination, a mouse, a keyboard, or other input modalities). As such, in operation 410, the received voice input may include a non-voice

Google Exhibit 1019
Page 00692

input component in addition to the utterance, such that an amount of input information may be maximized through usage of multi-modal inputs. For example, the user can use other input modalities to clarify a meaning of the utterance, provide additional information about the utterance, reduce a number of device interactions needed to make a given request, or otherwise maximize an amount of information provided in connection with a given utterance. For example, a multi-modal voice input may be used to efficiently interact with a map displayed on a navigation device (e.g., an utterance of "I only need to see this area" may be coupled with a stylus input that circles a given portion of a map, and the combination of inputs may result in a display of the circled portion of the map).

[073] Thus, for a given voice input, a decisional operation 420 may determine whether a non-voice component accompanies an utterance in the voice input. For example, in various implementations, the various input mechanisms associated with the voice user interface may be coupled to an Alternating Voice and Data (AVD) system. As a result, input data may alternate between voice and data, or the input data may be multiplexed together in a single stream, for example, to allocate processing resources where needed at a given moment.

[074] When decisional operation 420 determines that a non-voice input component has been received, the voice and non-voice components may be correlated in an operation 430. For example, the voice component may be processed using accurate speech recognition techniques, and the voice component may further be parsed and interpreted in light of context, shared knowledge, and available data sources, among other things, as described in greater detail above. As such, correlation operation 430 may generate additional information that can inform context, shared knowledge, or other information for processing the voice input. For instance, as discussed above, short-term shared knowledge may include awareness of various user interface states, which may be affected by non-voice components. Thus, correlation operation 430 may assist the voice user interface in developing one or more preliminary interpretations of the utterance, for example, by relating information in the utterance to information in the non-voice inputs, or vice versa. For example, a multi-modal input could include a user touching a stylus to one of a number of restaurant listings displayed on a touch-screen interface, while also providing an utterance relating to the touched listing (e.g., "Call this one."). In this example, correlation operation 430 can interpret "this one" to refer to the touched telephone listing, providing additional information for processing the voice input.

Google Exhibit 1019
Page 00693

[075] However, it will be apparent that voice inputs need not necessarily include non-voice components. As such, when the voice input includes only utterances, verbalizations, or other spoken components, decisional operation 420 may advance method 400 directly to an operation 440. For instance, as will be apparent from the further descriptions provided elsewhere in this specification, the natural language voice user interface may efficiently resolve many requests that include only voice. As such, a received voice input can include stand-alone utterances, such that a navigation device, vehicular system, mobile phone, or other device can be controlled in one step using voice. For example, using existing navigation systems or other navigation interfaces, a user must often take multiple steps to configure a displayed map, even for relatively simple tasks (e.g., multiple button and/or stylus inputs may be required to display a desired portion of a map). By contrast, using the natural language voice user interface, the user can simplify controlling a device using one-step voice control (e.g., a map may be controlled using utterances of "Move the map up," "Move the map Northeast," "Show me downtown Seattle," or other utterances that provide all information necessary to control a device in one-step), which may substantially reduce a number of device interactions needed to control the device.

[076] In various instances, (e.g., whether the input received in operation 410 includes an utterance, a non-voice component, or various combinations thereof), one or more queries and/or commands contained in the voice input may be identified in operation 440. Once the one or more preliminary interpretations of the multi-modal input have been generated, appropriate contextual information, shared knowledge, external knowledge sources, or other information may be used to identify the queries and/or the commands contained in the utterance. The information sources relied upon may be associated with various devices, systems, or other components of the integrated voice user interface environment. Thus, information associated with a navigation device (e.g., maps, points-of-interest, etc.) may be distributed for sharing with other devices in the environment (e.g., a personal digital assistant capable of rendering maps may display a map associated with a navigation device, or a mobile phone may dial a telephone number of a point-of-interest associated with the navigation device, etc.). Moreover, the information may be shared in a bi-directional manner, such that information sources associated various devices can be shared with other devices in the integrated environment.

[077] Information may be associated persistently (e.g., by storing all information on a server, an online data backup system, or other data repositories), or may be built dynamically as a

Google Exhibit 1019
Page 00694

function of processing multi-modal voice inputs. For example, a user may maintain an address book of contacts, e-mail addresses, street addresses, telephone numbers, or other information using a commercial service provider (e.g., GMail, Yahoo Mail, or another service provider capable of storing address books). When a user makes a request (e.g., via voice input, non-voice input, or multi-modal input) in which the service provider manages information needed to resolve the request, the information may be pulled and shared with a device that will be used to resolve the request. For example, a mobile phone may be integrated with the navigation voice user interface (e.g., a mobile phone having a GPS receiver), and the mobile phone may initially (e.g., at a first-time use) have no local contacts, addresses, telephone numbers, or other information. Thus, a user utterance of "I'm running late for dinner at Jane's house, call her to let her know I'm on the way" may result in a query being formulated to the service provider in order to retrieve an address and telephone number associated with Jane's house for the mobile phone to process accordingly. Moreover, information relating to being "late" may be processed to generate a route that avoids traffic, uses highways, or otherwise reduces an amount of travel time.

[078] A variety of cognitive models, contextual models, user-specific models, or other models may be used to identify the query and/or command in operation 440. For example, a given input may include information relating to one or more contextual domains, one or more of which may be invoked to interpret and/or infer certain keywords, concepts, or other information contained in the input. Moreover, short-term and long-term shared knowledge about a user's behavior and preferences may be used in a hybrid recognition model, for example, relating to navigation queries and/or commands. The hybrid recognition model may be used to generate probable or likely interpretations of an utterance using a combination of semantic analysis and contextual inferences. For example, certain syllables, words, phrases, requests, queries, commands, or other identifiable aspects of an utterance may be identified as being more likely to occur based on a given contextual history. The contextual history can be based on the short-term or the long-term shared knowledge about the user, previous utterances in a current conversation, common requests in a given environment (e.g., a navigation device may be used most often for displaying maps, calculating routes, and identifying locations), or other information. Thus, the hybrid recognition model for operation 440 may include various processes for analyzing semantic

Google Exhibit 1019
Page 00695

patterns to resolve what was said by an utterance, in addition to various processes for relying on contextual history to resolve what was meant by the utterance.

[079] In addition, the hybrid recognition model may be used in conjunction with, or independently of, a peer to peer recognition model. For example, contextual histories may include various preferences or user characteristics, in addition to providing a basis for inferring information about a user based on patterns of usage or behavior, among others. As a result, the recognition models may include additional awareness relating to global usage patterns, preferences, or other characteristics of peer users on a global basis. For example, certain keywords, concepts, queries, commands, or other aspects of a contextual framework may be commonly employed by all users within a context. In another example, users of certain demographics may share common jargon, slang, or other semantic speech patterns. As a result, operation 440 may utilize various recognition models, which consider context and semantics in various dimensions, to identify queries or command. For example, in addition to information generally available within a given environment or context, a voice input may be recognized using context and semantics based on short-term or long-term behavior or preferences for a specific user, global users, peer users, or other meaningful user abstractions.

[080] Accordingly, upon identifying a suitable query and/or command contained in the utterance, a suitable response may be generated, or a task may be performed in an operation 450. As will be apparent from the descriptions already provided, the generated response and/or the performed task may be highly dependent on context and information contained in a given voice input. As a result, when certain voice inputs may be unclear, ambiguous, or otherwise unclear, the generated response may prompt a user for an additional input, and the response may be framed in a way favorable for recognition (e.g., an utterance of "Where is Joe's Pizza" provided while a user is in New York City may return too many results to be meaningful, such that results may be weighed and displayed for a user to select a correct one of the results).

[081] Furthermore, voice inputs may be used to perform compound requests, which could otherwise be impossible to perform using a single manual input. For example, a single voice-based input may include a compound map control request, such as "Show me downtown Seattle." As such, operation 450 may perform tasks of retrieving a map of Seattle and automatically zooming in on a downtown area of the Seattle map. Further, one or more responses or other outputs may be generated, such as suggesting a route to a frequent destination,

Google Exhibit 1019
Page 00696

identifying possible points-of-interest to the user, or searching for traffic or event notifications, among other things. Many other variations will be apparent, including characteristics of the received inputs (e.g., utterances, non-voice inputs, etc.), the requested queries or commands, the generated responses, and the performed tasks, among other things. As such, it will be apparent that the method 400 illustrated herein may enable users to request many different navigation related tasks verbally, non-verbally, or various combinations thereof, such that the users can request various kinds of information or tasks available in the environment (e.g., relating to various devices, information sources, etc.). Thus, method 400 may operate in an environment in which the natural language voice user interface has been associated with one or more devices that provide navigation services.

[082] According to various aspects of the invention, Figure 5 illustrates a flow diagram of an exemplary method for calculating navigation routes from voice destination entries provided to a navigation device having a conversational, natural language voice user interface. As previously described, the voice user interface may calculate routes, provide dynamic data-driven directions to a destination, provide dynamic routing to a destination, perform post-processing of full or partial destination entries, or otherwise provide various voice navigation services. Thus, as illustrated in Figure 5, the method for calculating routes may support successive refinement of a voice destination entry, wherein context, agent adaptation, and shared knowledge, among other things, can help a user to narrow down a final destination using voice commands, multi-modal commands, or various combinations thereof. However, though Figure 5 specifically illustrates an exemplary implementation of successive refinement for voice destination entry, it will be apparent that the techniques described herein can be applied in performing various tasks in which generalized approximations can suitably be refined through successive voice or multi-modal commands to narrow down information sought by a user in various domains, contexts, applications, devices, or other components that employ the techniques described herein. As a result, successive refinement can be implemented in various domains that enable a user to "drill down" to a specific piece of information or data through specifying more and more specific criteria about the information sought.

[083] As illustrated in Figure 5, for example, successive refinement of a final destination may includes instances in which a user specifies the destination in a single utterance, over a plurality of utterances (e.g., which sequentially narrow down approximations of the destination), as well

Google Exhibit 1019
Page 00697

as instances in which a system identifies the destination (e.g., through generating inferences, using context or shared knowledge, etc.). For example, a given approximation of a final destination may be associated with a finite number of possible destinations, and the system may be able to unambiguously identify the final destination from among the finite possible destinations by generating inferences, by relying on context, shared knowledge, and other information, or by other ways, as will be apparent. Moreover, in another example, successively refining the destination may be modeled after patterns of human interaction between passengers in a taxicab, or other similar situations in which a route or a destination may be narrowed down or otherwise refined over a course of interaction. For example, passengers in taxicabs sometimes specify a general approximation of a destination (e.g., a passenger picked up in Arlington, Virginia may state, "I'm going to the Foggy Bottom area of D.C."), which may result in a cab driver beginning to drive along a preferred route to the approximated destination (e.g., the driver may know to head towards Rock Creek Parkway when driving from Arlington to Foggy Bottom). While en route to the approximated destination, the passenger and/or the driver may cooperate to refine the final destination through one or more subsequent interactions (e.g., the cab driver may ask, "Where in Foggy Bottom are you headed?", and the passenger may specify "McFadden's," which may provide sufficient information for the driver to determine a route to the destination).

[084] However, it will be apparent that no two sequences of interactions will necessarily be alike. For example, the passenger may provide additional information relating to the final destination without being prompted by the driver, or the driver may not begin driving until the final destination has been identified (e.g., when several routes may be available, and the final destination has an impact on which route will be selected). As such, cooperative, conversational interactions may have a general goal of calculating a route to a final destination, yet the interactions may occur in many different ways, where a given interaction will likely be highly dependent on context, shared knowledge, agent adaptation, points of origin and/or presence, user preferences, dynamic data (e.g., relating to traffic, events, external systems, etc.), among many other factors.

[085] Accordingly, as further described herein, the method illustrated in Figure 5 may enable voice destination entry through successive refinement using the aforementioned factors, among other things, to inform route calculation based on full or partial destination inputs. For instance,

Google Exhibit 1019
Page 00698

in similar fashion as described in the above cab driver example, the voice user interface (e.g., the cab driver) may have knowledge that certain routes may be preferable to reach a certain destination from a current point of presence (e.g., Arlington to Foggy Bottom via Rock Creek Parkway). Further, as will be described in greater detail in Figure 6, a search space associated with subsequent interactions may be constrained according to a current route, context, or other knowledge (e.g., locations named "McFadden's" that fall outside a certain proximity of Foggy Bottom may be excluded).

[086] Thus, in an operation 510, a user may provide a full or partial destination input using free form natural language, for example, including voice commands and/or multi-modal commands (e.g., an input may include an utterance of "I'm going here," coupled with a touched point on a display). The full or partial destination may be specified in various ways, including by specific address, place name, person's name, business name, neighborhood, city, region, or various other ways (e.g., a voice destination entry may be provided in an exploratory manner, such as when a user wants to visit a museum, but has yet to decide which one to visit). Furthermore, because successive refinement tolerates incomplete or partial addresses, voice destination inputs may be provided in a way that uses the successive refinement in reverse, for example, by narrowly specifying a desired destination in order to calculate a route to a broader vicinity of the desired destination (e.g., a route to South Philadelphia could be identified through an utterance of "Pat's Steaks in Philly," even though Pat's Steaks may not be a user's actual destination).

[087] The voice destination input may be parsed or otherwise analyzed using one or more dynamically adaptable recognition grammars, for example, as described above in reference to Figure 3. For example, recognition grammars may be loaded, generated, extended, pruned, or otherwise adapted based on various factors, including a proximity to a user's point of presence (e.g., as the user moves from one area to another, the recognition grammar may be optimized based on a current location, a direction of travel, temporal constraints, etc.), a contextual history (e.g., as the user interacts with the voice user interface, the grammar may adapt based on dictionaries, keywords, concepts, or other information associated with other contexts, domains, devices, applications, etc.), or other factors, as will be apparent. As such, an operation 520 may include generating one or more interpretations of the voice destination input, which may be analyzed using various data sources in order to generate an N-best list of possible destinations (e.g., a navigation agent may query a directory, a voice search engine, or other components to

Google Exhibit 1019
Page 00699

identify one or more destinations that at least partially match criteria contained in the voice destination input).

[088] The generated list of possible destinations may be post-processed in an operation 530 in order to assign weights or ranks to one or more of the entries in the N-best list. The post-processing may include analyzing the destination list generated in operation 520 according to various factors in order to determine a most likely intended destination from a full or partial voice destination input. For example, post-processing operation 530 may rank or weigh possible destinations according to shared knowledge about the user, domain-specific knowledge, dialogue history, or other factors. Furthermore, the post-processing operation 530 may analyze the full or partial destination input in order to identify an address to which a route can be calculated, for example, by resolving a closest address that makes "sense" relative to the input destination. For example, a user may specify a partial destination that identifies a broad and approximated area (e.g., "Take me to Massachusetts"), and depending on a user's current location, direction of travel, preferences, or other information, post-processing operation 530 may select an address makes sense for calculating a route (e.g., an address in Cape Cod may be selected for a user having relatives that live in Cape Cod, whereas an address in Boston may be selected for a user who may be traveling to various popular sightseeing areas, etc.).

[089] As a result, the weighed list of N-best destinations may be evaluated in an operation 540 to determine a suitably identifiable destination exists in the list. For example, a full or partial voice destination entry may be ambiguous, or certain criteria or keywords in a voice destination entry may be unrecognizable, such that a highest ranked destination in the weighted list does not exceed a minimal confidence level needed to identify a destination. For instance, a user located in Oklahoma may utter a partial destination, as in "I'm heading to Washington," and the decisional operation 540 may return a negative indication when the voice user interface cannot disambiguate between Washington state, Washington, D.C., Washington University in Saint Louis, and a town of Washington located slightly south of Oklahoma City. In another example, a user originating in Saint Louis may provide a voice destination entry of "Take me to Springfield," which could result in an unidentifiable destination even though multiple destinations may satisfy the minimal confidence level. In this example, "Springfield" may be unidentifiable as a destination because Springfield, Illinois and Springfield, Missouri both reside within reasonable distances of Saint Louis, yet directions of travel to either destination may be

Google Exhibit 1019
Page 00700

entirely opposite (e.g., a route to the Illinois Springfield includes traveling north on Interstate-55, whereas a route to the Missouri Springfield includes traveling southwest on Interstate-44). Thus, to avoid routing the user in a direction opposite from an intended destination, processing may instead branch to an operation 550 to generate a prompt for resolving the destination.

[090] For example, the prompt may ask the user a question or provide information that can clarify an intended meaning of the received voice destination input. For instance, in the Washington example given above, a multi-modal user prompt may display the four possible destinations on a display device (i.e., Washington state, Washington, D.C., Washington University, and Washington, Oklahoma). The multi-modal prompt may further include system-generated speech, for example, stating, "I found several possible Washingtons, did you mean one of these or something else?" As a result, processing then proceeds back to operation 510, in which the user can disambiguate the intended destination through another input (e.g., by touching one of the displayed destinations when the display device includes a touch-screen, or by verbalizing one of the destinations, or by providing additional information indicating that none of the displayed destinations were intended). Thereafter, subsequent N-best lists may be more likely to result in an identifiable destination, as an original voice destination input can be interpreted in view of a dialogue history.

[091] Alternatively, when at least one of the possible destinations can be identified unambiguously, while meeting the minimal confidence level, a positive indication may result in processing proceeding to another decisional operation 560, which controls how a route will be calculated to the identified destination (e.g., a highest ranking entry in the weighted N-best list of destinations). The decisional operation 560 may determine whether the identified destination provides a full or otherwise complete destination. For example, through successive refinement, a partial voice destination entry may result in an identifiable, yet incomplete, destination. As such, an operation 570 may calculate a preliminary route to the identified partial destination. Thereafter, subsequent processing returns to operation 510, where the voice user interface may await an additional voice destination input that refines the partial destination. Further, in various implementations, one or more system prompts may be generated to request the additional voice destination inputs (e.g., as the user approaches the partial destination and additional information will be needed to provide further routing, or as the user approaches a point in the preliminary

Google Exhibit 1019
Page 00701

route where distinct routes may be taken to different topographical subtiles or points within the partial destination, etc.).

[092] As such, the processing operations 510-570 may be iteratively repeated until the final destination can be identified through successive refinement using one or more multi-modal voice destination entries. As a result, when a suitable final destination has been identified, operation 560 may be followed by an operation 580 that completes the route to the final destination. For example, the route may be completed by extrapolating the route to the partial destination into a complete route to the final destination. Further, it will be apparent that voice destination entries may be successively refined into a final destination en route (e.g., the final destination may be successively refined as a user proceeds along a preliminary route), in advance (e.g., the user may choose to drive upon a complete route to a final destination being identified), or in other ways.

[093] Moreover, as described in greater detail herein, the calculated route may be dynamically adjusted or rerouted based on subsequent inputs, generated inferences, dynamic data, or various other sources of information (e.g., inferences may be generated, which may result in dynamic routing, in response to dynamic data relating to traffic conditions, detours, events, weather conditions, or user preferences, or other factors). Thus, Figure 6 illustrates a flow diagram of an exemplary method for providing voice services based on a current navigation route according to various aspects of the invention. The voice user interface may use the method illustrated in Figure 6 to provide the voice services in a dynamic, data-driven manner, wherein the voice services may be provided in a customized manner, for example, based on context, shared knowledge about a user, or other sources of information.

[094] Accordingly, providing voice navigation services based on a current route includes identifying the current navigation route in an operation 610. The route may be identified from a preliminary route or a complete route, as calculated according to the techniques described above in reference to Figure 5. Thus, current routing information may be made available to various system components in order to process events or voice inputs, generate responses, formulate queries, resolve commands, or perform other processing services available in a voice navigation services environment. As such, the current routing information may provide further context relevant to interpreting voice inputs, taking action towards resolving the inputs, or responding to events, among other things. Thus, though not illustrated in Figure 6, various other sources of contextual information may be utilized to provide voice navigation services in view of the

Google Exhibit 1019
Page 00702

current route. For example, user preferences may limit places to which a vehicle can travel, or for how long the vehicle can travel, among other things (e.g., for keeping a delivery or passenger route on a regular route, or for placing parental controls on teenage drivers, etc.). In another example, the voice user interface may have awareness of one or more external systems or devices, sources of dynamic data, or other information, which can be utilized to determine how to best provide the voice navigation services.

[095] Thus, upon identifying the current route in operation 610, along with other appropriate information relating to context, shared knowledge, or otherwise, a decisional operation 620 may determine whether the identified information results in an event being generated and/or detected. When no event can be generated and/or detected, processing proceeds to a decisional operation 630, which determines whether a voice input has been received from the user. As such, the method illustrated herein may loop through decisional operations 620 and 630 until one or more of an event has been detected and/or a voice input has been received.

[096] One type of event that can be detected and/or generated in decisional operation 620 may relate to location dependent advertisements for navigation systems (e.g., as generated by a local or remote advertising engine, or via a data channel, or in other ways). For example, in various implementations, the advertising events may be generated and/or detected using techniques described in co-pending U.S. Patent Application 11/671,526, entitled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input," filed February 6, 2007, the disclosure of which is hereby incorporated by reference in its entirety. For instance, as will be apparent, a navigation system will necessarily include one or more systems or mechanisms for determining the navigation system's current location (e.g., a global positioning system, a radio frequency identification system, a system that determines location based on a distance to an identifiable wireless tower or access point, etc.). Further, the natural language voice user interface may be able to determine location in a number of other ways, in addition to or apart from the location detection system (e.g., an adaptable prompt may affirmatively request the location information from a user, for example, upon failure of the location detection system, or other techniques utilizing shared knowledge or context may be used, as will be apparent).

[097] Thus, in one example, operation 620 may include detecting and/or generating events relating to a location dependent advertisement. A marketer or other entity seeking to reach a

Google Exhibit 1019
Page 00703

target audience may simply scatter radio frequency identifiers (RFIDs) along a roadside, or another place in which a system may detect the RFIDs. The marketer or other entity may then broadcast one or more advertisements via a data channel associated with the RFIDs, such that the navigation system may trigger an event when within a suitable proximity of the RFIDs. Thus, processing would proceed to an operation 570, in which information associated with the event may be filtered according to the current routing information or other contextual parameters. As applied to this example, the filtering operation 650 may include monitoring the data channel to analyze information associated with the detected RFID, thus receiving the advertisements broadcasted over the data channel. Thus, an advertisement including speech, for example, may be analyzed in a similar manner to how user speech would be analyzed (e.g., maintaining awareness that the speech relates to an advertisement).

[098] Accordingly, the advertisement event data may be filtered in view of the current route, personal information (e.g., previous queries, preferences, history, etc.), or other factors to determine what action should be taken in an operation 660. For example, the advertisement may be associated with a local restaurant seeking to reach personal navigation devices that may not necessarily have a complete business database. Therefore, in one example, filtering operation 650 could discard the advertisement when shared knowledge indicates that the user prefers to reach destinations as fast as possible, without making unnecessary stops. In this case, operation 660 may effectively be a null operation, or may simply include displaying an image associated with the advertisement to allow the user to make an independent choice as to whether or not to stop. However, in a distinct example, the user may have provided a previous voice input indicating a desire to stop driving and eat lunch at some point along the route. Thus, in this example, operation 660 may include a spoken output that brings the restaurant to the user's attention, possibly asking the user whether the route should be temporarily diverted to stop for lunch at the restaurant.

[099] However, it should be appreciated that the above described RFID-based advertisement provides but one example of a possible advertising event that may be detected in operation 620. For instance, advertisements may be uploaded to a server by one or more advertising partners, wherein the uploaded advertisements may be associated with metadata or other descriptive information that identifies a target audience, location-dependent information, or other criteria. In another example, a plurality of advertisements may be stored locally at the voice user interface,

Google Exhibit 1019
Page 00704

and an inferencing engine may determine appropriate circumstances in which an event should be generated to deliver one or more of the advertisements to a user. As a result, it will be apparent that advertising events may be generated in a number of ways, and may be generated and/or detected locally, remotely, by detecting RFIDs, or in other ways.

[0100] Further, as advertisements get delivered to users (e.g., in operation 660), the users' subsequent interaction with the delivered advertisements may be tracked. In this way, affinity based models may be generated, for example, to ensure that promotions or advertisements will be delivered to a likely target audience. Thus, an event relating to a given advertisement may be generated and/or detected in operation 620 when shared knowledge about a user's behavior, preferences, or other characteristics match one or more criteria associated with peer-to-peer affinities associated with the advertisement. For example, a marketer may be offering promotional coupons intended for a specific demographic audience, and the peer-to-peer affinities may be used to ensure that the coupons will only be provided to users likely to be in the intended audience (e.g., to avoid providing an entire market, including unlikely customers, with the coupons). Thus, when the user falls within the specified demographic audience, one or more events may be generated on a server, locally by an advertising engine, or in other ways to deliver one or more of the promotional coupons to the user.

[0101] In other examples, the advertising model may include mobile pay-per-use systems (e.g., network access may be provided through a wireless access point on a pay-per-use basis), peer-to-peer local guides or recommendations (e.g., affinities may indicate that various categories of peer users may be interested in similar points-of-interests, events, local restaurants, stores, activities, tourist attractions, or other aspects of a local area). Additionally, as described in greater detail above in reference to Figure 3, various aspects of the advertising model, such as the local guides and recommendations, may be generated according to a mapping applied to various topological domains. For example, local guides or recommendations may be dependent on topological characteristics (e.g., different guides for various cities, communities, localities, etc.). As such, the local guides, recommendations, or other aspects of the advertising model may be associated with a topological taxonomy based on geographical chunking. As a result, various advertising events may be generated and/or detected according to physical proximities, temporal proximities, directional proximities, civil organization proximities, or various combinations

Google Exhibit 1019
Page 00705

thereof (e.g., peer-to-peer recommendations may be varied depending on a user's location, direction of travel, temporal constraints, etc.).

[0102] In addition to the advertising related events that can be generated and/or detected in operation 620, the natural language voice user interface may generate additional events through awareness of context, shared knowledge about a user, external systems and devices, or other information. For example, as discussed above in reference to Figure 2, the voice user interface may implemented within an agent-based architecture, in which one or more of a plurality of agents may include an inference engine. However, it will be apparent that the inference engine may be arranged within the voice user interface in various other ways. For example, the voice user interface may include the inference engine within a conversational language processor, where in such implementations, the inference engine may generate, detect, and distribute events to various components of the voice user interface. In another example, a managing inference engine may coordinate event generation and/or detection among the inference engines associated with the agents. As such, it will be apparent that the voice user interface may include various suitable arrangements of one or more inference engines, such that events and inferences can be detected, generated, and distributed to various system components as they may arise.

[0103] Thus, the one or more inference engines may utilize awareness of context, shared knowledge about a user, dynamic data sources, data and services associated with external or devices, among other information to generate and/or detect the events or other inferences identified in operation 620. When the events or other inferences occur when a current navigation route exists, the events or other inferences may be filtered in operation 650 prior to being a response being generated and/or a task being performed in operation 660.

[0104] For example, the inference engine may utilize personal information, such as a user frequently visiting a particular person in the user's address book, to generate an event that suggests rerouting the current route. Thus, in this example, the inference engine may generate the event (which may be detected in operation 620) when the current route or current location information reflects a certain proximity to an address associated with the person in the address book. Operation 650 could then filter the event by determining possible responses or courses of action for handling the event in a manner consistent with the current route. For example, information associated with the current route may indicate that the user was out running errands and may now be heading home. Further, a calendar in the user's address book may indicate that

Google Exhibit 1019
Page 00706

the user has an appointment later in the day, but that no calendar entries exist for a few hours. Thus, an exemplary response generated in operation 660 may include a speech-based output of, "You are close to Jane's house, would you like to stop by?"

[0105] Although the above-provided example illustrates an event generated using personal information, it will be apparent that various available information sources may be utilized to generate and/or detect events. For example, events may be based on transient or dynamic data relating to communities, traffic, weather, and/or many other sources of data (e.g., a generated event may result in an output of "Tonight's soccer game is being played at a field near your house, do you want to go?", in which the soccer game may be a local community event).

[0106] In yet another example, various types of external system awareness may trigger events and resulting voice responses from the voice user interface. For example, the voice user interface may be coupled to various vehicular or telematics systems, providing awareness over a state of such systems. Thus, an event may be generated when the vehicular system indicates that the vehicle will soon run out of gas, resulting in a voice response prompting a user to stop for gas. Further, the external system awareness may be filtered according to the current routing information (e.g., in conjunction with the inference engine), wherein a calculation may be made indicating that a level of gas will be sufficient to get to a next gas station along the route, but that the level of gas will not be sufficient to get to a second gas station along the route. As a result, the route-based filtering operation 650 may result in a response being generated in operation 660 that provides a voice warning that the user must stop for gas at the next gas station.

[0107] Returning to operation 630, user-provided voice inputs may also be processed in view of current routing information. As previously discussed, interpretations of what was said in a voice input (e.g., content of an utterance) may be based on various phonetic models, dictionaries, context histories, dialogue histories, and other information that can form a dynamic recognition grammar. Moreover, context, shared knowledge, and available data and services, among other information, may be used to interpret what meant by the voice input (e.g., an intent of the utterance). As a result, using the various techniques described herein and in the incorporated co-pending U.S. Patent Applications, a query, command, or other request contained in the utterance may be identified in an operation 640. When a current route exists, a domain of possible tasks, commands, query answers, responses, and/or other system actions may be filtered according to the current route. For example, a voice input of "Where's the closest bathroom" may be

Google Exhibit 1019
Page 00707

analyzed in view of the current route. Various proximities (e.g., a direction of travel, physical distances, temporal distances, etc.) may then be utilized to determine what bathrooms may be appropriate in view of the current route. Thus, when a route has the user traveling a long distance, for example, a bathroom at a rest area twenty miles ahead may be favored over a bathroom at a restaurant ten miles off a highway (e.g., to avoid diverting the user from the current route).

[0108] Furthermore, in many instances, an agent or other processing component may need to formulate a plurality of queries to various different data sources in order to resolve a request. As such, operation 660 for performing a task may determine what inferences, metadata, or other information may be needed to formulate queries to the data sources in order to resolve the request. As a result, voice requests relating to navigation may include information that may not necessarily be linked to navigation, or that may not necessarily be in a database or other source of navigation information. Even so, by being coupled with various available sources of information, complex or compound queries may be formulated to retrieve information from various data sources in order to resolve a given request. Furthermore, a given utterance may include multiple requests, queries, or commands, such that a given response or task performed in response to the utterance may include interactions with various different system components. Thus, in an illustrative example, an utterance of "Find the 7-11 closest to the Woodland Park Zoo" may include a first query to identify a location of Woodland Park Zoo, a second query to find one or more 7-11's within a certain radius of the Woodland Park Zoo location, a third query to determine a current location of the user, and a fourth query that invokes an inference engine to determine which 7-11 may be "closest" based on various user proximities. Thus, supposing that two 7-11's exist equidistantly from the Woodland Park Zoo, one in a direction east of the Zoo and one in a direction west of the Zoo, the user's current location may be used to determine which 7-11 to select; by contrast, a selection based purely on physical proximity to the Zoo could result in ambiguity.

[0109] Although the above-provided descriptions highlight various specific examples and illustrations of how the natural language voice user interface may provide voice navigation services, it will be understood that the descriptions, examples, and modes of operation provided above should be regarded as exemplary only. The natural language voice user interface may be highly tolerant of variations in how humans interact verbally, and may utilize context, shared

Google Exhibit 1019
Page 00708

knowledge, and available data and systems to effectively interact with any given user in a conversational manner that advances mutual goals. As a result, interactions between the user and the voice user interface may be regarded as unique occurrences, as users may speak free form requests relating to various available data sources, applications, or devices in a given interaction. Thus, the voice user interface may dynamically adapt, learn, and otherwise refine an operational framework to provide optimal services in the voice navigation services environment.

[0110] Implementations of the invention may be made in hardware, firmware, software, or various combinations thereof. The invention may also be implemented as instructions stored on a machine-readable medium, which may be read and executed by one or more processors. A machine-readable medium may include a mechanism for storing or transmitting information in a form readable by a machine (e.g., a computing device). For example, a machine-readable storage medium may include read only memory, random access memory, magnetic disk storage media, optical storage media, flash memory devices, and others, and a machine-readable transmission media may include forms of propagated signals, such as carrier waves, infrared signals, digital signals, and others. Further, firmware, software, routines, or instructions may be described in the above disclosure in terms of specific exemplary aspects and implementations of the invention, and performing certain actions. However, those skilled in the art will recognize that such descriptions are merely for convenience and that such actions in fact result from computing devices, processors, controllers, or other devices executing the firmware, software, routines, or instructions.

[0111] Aspects and implementations may be described as including a particular feature, structure, or characteristic, but every aspect or implementation may not necessarily include the particular feature, structure, or characteristic. Further, when a particular feature, structure, or characteristic is described in connection with an aspect or implementation, it will be understood that one skilled in the art may be able to effect such feature, structure, or characteristic in connection with other aspects or implementations, whether or not explicitly described. Thus, various changes and modifications may be made to the provided description without departing from the scope or spirit of the invention. As such, the specification and drawings should be regarded as exemplary only, and the scope of the invention to be determined solely by the appended claims.

Google Exhibit 1019
Page 00709

CLAIMS

What is claimed is:

1. A computer-implemented method for providing a natural language voice user interface for a navigation device, comprising:

receiving a navigation-related voice-based input at an input mechanism associated with the navigation device, the voice-based input including at least a natural language utterance provided by a user;

generating one or more preliminary interpretations of the natural language utterance using a dynamic recognition grammar associated with a speech recognition engine;

analyzing the preliminary interpretations using a conversational language processor, the conversational language processor using shared knowledge and information associated with a navigation context to determine a probable interpretation of the natural language utterance in the navigation context; and

communicating the probable interpretation of the natural language utterance to a navigation agent associated with the navigation context, the navigation agent operable to:

identify one or more navigation-related requests contained in the voice-based input based on the probable interpretation of the natural language utterance; and

resolve the requests using information associated with one or more of a plurality of information sources, the plurality of information sources including at least a navigation-specific information source.

2. The method of claim 1, the navigation agent further operable to determine that at least one of the requests includes an approximation of requested information, wherein the conversational language processor manages a dialogue with the user to resolve the approximated information, the managed dialogue including one or more interactions with the user that successively refine the approximated information.

3. The method of claim 2, each of the interactions including one or more of prompting the user to refine the approximated information or receiving a subsequent input from the user that refines the approximated information.

Google Exhibit 1019
Page 00710

4. The method of claim 1, the navigation-related requests including a request for a route to a full or partial address, the navigation agent operable to resolve the request for the route by:

selecting a destination for the route, the selected destination having an address that best corresponds to the full or partial address;

calculating a route from a current location of the user to the selected destination; and

generating directions to guide the user on the calculated route to the selected destination, the generated directions dynamically driven by data from a plurality of information sources.

5. The method of claim 4, the dynamic data-driven directions including information about destinations, points of interest, traffic, parking, weather, events, or other information relevant to the calculated route.

6. The method of claim 4, further comprising:

receiving a subsequent voice-based input at the input mechanism associated with the navigation device, the subsequent voice-based input including at least one request provided by the user; and

invoking at least one of a plurality of domain agents to resolve the request included in the subsequent voice-based input, the at least one domain agent operable to filter results of the request according to the calculated route.

7. The method of claim 1, the dynamic recognition grammar including information associated with one or more topological domains.

8. The method of claim 7, the topological domains including physical, temporal, directional, and civil organizational proximities with respect to a current location of the user.

9. The method of claim 1, the conversational language processor coupled to an inferencing engine operable to generate one or more inferences, wherein the conversational language processor uses the generated inferences to determine the probable interpretation of the natural language utterance.

Google Exhibit 1019
Page 00711

10. The method of claim 9, wherein the conversational language processor further uses the generated inferences to suggest one or more services available in the navigation context to the user.

11. The method of claim 1, the navigation-related requests including a multi-modal request to control a map display, the navigation agent operable to resolve the request to control the map display by:

associating a non-voice component of the voice-based input with the probable interpretation of the utterance, the non-voice component identifying a portion of the map display; and

issuing a command to control the identified portion of the map display in accordance with the probable interpretation of the utterance.

12. The method of claim 1, wherein the plurality of information sources include the navigation-specific information source, the shared knowledge, the information associated with the navigation context, and one or more information sources relating to maps, destinations, directories, points of interest, traffic, parking, weather, events, user address books, user devices and systems, a search engine, and a plurality of domain agents.

13. The method of claim 12, the shared knowledge including dialogue history information, request history information, user interface state information, short-term user profile information, long-term user profile information, peer user profile information, and user location information.

14. A system for providing a natural language voice user interface for a navigation device, comprising:

an input mechanism that receives a navigation-related voice-based input, the voice-based input including at least a natural language utterance provided by a user;

a speech recognition engine having a dynamic recognition grammar associated therewith, the speech recognition engine operable to generate one or more preliminary interpretations of the natural language utterance using the dynamic recognition grammar;

Google Exhibit 1019
Page 00712

a plurality of domain agents associated with respective contexts, the plurality of domain agents including at least a navigation agent associated with a navigation context; and

a conversational language processor operable to:

analyze the preliminary interpretations using shared knowledge and information associated with the navigation context to determine a probable interpretation of the natural language utterance in the navigation context; and

communicate the probable interpretation of the natural language utterance to the navigation agent associated with the navigation context, the navigation agent operable to identify one or more navigation-related requests contained in the voice-based input based on the probable interpretation of the natural language utterance and resolve the requests using information associated with one or more of a plurality of information sources, the plurality of information sources including at least a navigation-specific information source.

15. The system of claim 14, the navigation agent further operable to determine that at least one of the requests includes an approximation of requested information; and

the conversational language processor further operable to manage a dialogue with the user to resolve the approximated information, the managed dialogue including one or more interactions with the user that successively refine the approximated information.

16. The system of claim 15, each of the interactions including one or more of prompting the user to refine the approximated information or receiving a subsequent input from the user that refines the approximated information.

17. The system of claim 14, the navigation-related requests including a request for a route to a full or partial address, the navigation agent operable to resolve the request for the route by:

selecting a destination for the route, the selected destination having an address that best corresponds to the full or partial address;

calculating a route from a current location of the user to the selected destination; and

generating directions to guide the user on the calculated route to the selected destination, the generated directions dynamically driven by data from a plurality of information sources.

Google Exhibit 1019
Page 00713

18. The system of claim 17, the dynamic data-driven directions including information about destinations, points of interest, traffic, parking, weather, events, or other information relevant to the calculated route.

19. The system of claim 17, the one or more input mechanisms further receiving a subsequent voice-based input, the subsequent voice-based input including at least one request provided by the user, and the conversational language processor further operable to:

invoking at least one of a plurality of domain agents to resolve the request included in the subsequent voice-based input, the at least one domain agent operable to filter results of the request according to the calculated route.

20. The system of claim 14, the dynamic recognition grammar including information associated with one or more topological domains.

21. The system of claim 20, the topological domains including physical, temporal, directional, and civil organizational proximities with respect to a current location of the user.

22. The system of claim 14, further comprising an inferencing engine coupled to the conversational language processor that generates one or more inferences, wherein the conversational language processor uses the generated inferences to determine the probable interpretation of the natural language utterance.

23. The system of claim 22, further comprising an inferencing engine coupled to the conversational language processor, wherein the conversational language processor further uses the generated inferences to suggest one or more services available in the navigation context to the user.

24. The system of claim 14, the navigation-related requests including a multi-modal request to control a map display, the navigation agent operable to resolve the request to control the map display by:

Google Exhibit 1019
Page 00714

associating a non-voice component of the voice-based input with the probable interpretation of the utterance, the non-voice component identifying a portion of the map display; and

issuing a command to control the identified portion of the map display in accordance with the intended meaning of the utterance.

25. The system of claim 14, wherein the plurality of information sources include the navigation-specific information source, the shared knowledge, the information associated with the navigation context, and one or more information sources relating to maps, destinations, directories, points of interest, traffic, parking, weather, events, user address books, user devices and systems, a search engine, and a plurality of domain agents.

26. The system of claim 25, the shared knowledge including dialogue history information, request history information, user interface state information, short-term user profile information, long-term user profile information, peer user profile information, and user location information.

27. A method for providing a natural language voice user interface in a voice navigation services environment, comprising:

receiving a voice-based destination input provided by a user, the voice-based destination input including at least a natural language utterance;

generating one or more preliminary interpretations of the natural language utterance using a dynamic recognition grammar associated with a speech recognition engine;

analyzing the preliminary interpretations using a conversational language processor, the conversational language processor using shared knowledge and information associated with a navigation context to determine an intended destination provided in the natural language utterance, the intended destination including an approximation of a destination; and

communicating the intended destination to a navigation agent, the navigation agent operable to provide a route to the intended destination by:

selecting a preliminary destination for the route, the selected preliminary destination having an address that best corresponds to the approximated destination;

Google Exhibit 1019
Page 00715

calculating a route from a current location of the user to the selected preliminary destination; and

participating in a dialogue with the user to resolve a final destination for the calculated route, the dialogue includes one or more interactions with the user that successively refine the approximated destination until the final destination has been resolved.

28. The method of claim 27, each of the interactions including one or more of prompting the user to refine the approximated destination or receiving a subsequent input from the user that refines the approximated information.

29. The method of claim 27, the navigation agent further operable to dynamically recalculate the route in response to the successive refinements of the approximated destination.

30. The method of claim 27, the final destination being resolvable substantially later in time with respect to an initial calculation of the route.

31. The method of claim 27, the navigation agent operable to select the preliminary destination by identifying one or more addresses possibly corresponding to the approximated destination and selecting a highest ranked one of the identified addresses as the preliminary destination.

32. The method of claim 27, the identified addresses ranked according to a proximity to the current location of the user or the approximated destination.

33. The method of claim 32, the proximity to the based on one or more topological domains, including one or more of physical, temporal, directional, and civil organizational proximities with respect to the current location of the user or the approximated destination.

Google Exhibit 1019
Page 00716

34. The method of claim 27, the navigation agent further operable to generate directions to guide the user on the calculated route, the generated directions dynamically driven by data from a plurality of information sources.

35. The method of claim 34, the plurality of information sources including information relating to destinations, points of interest, traffic, parking, weather, events, or other information relevant to the calculated route.

36. The method of claim 27, the shared knowledge including dialogue history information, request history information, user interface state information, short-term user profile information, long-term user profile information, peer user profile information, and user location information.

37. A method for providing a natural language voice user interface in a voice navigation services environment, comprising:

receiving a navigation-related voice-based input provided by a user, the voice-based input including at least a natural language utterance;

identifying a current location of a user using a location detection system;

determining one or more topological domains associated with the current location of the user;

generating a dynamic recognition grammar that includes grammar information associated with the determined topological domains; and

using the dynamic recognition grammar to generate one or more interpretations of the natural language utterance.

38. The method of claim 37, the dynamic recognition grammar organizing the grammar information according to geographic chunks associated with the topological domains.

39. The method of claim 38, the geographic chunks including physical proximities defined by a distance from the current location of the user.

Google Exhibit 1019
Page 00717

40. The method of claim 38, the geographic chunks including temporal proximities defined by a time of travel from the current location of the user.

41. The method of claim 38, the geographic chunks including directional proximities defined by a directional vector of travel by the user.

42. The method of claim 38, the geographic chunks including civil organizational proximities defined by continents, countries, regions, states, cities, localities, neighborhoods, and communities.

43. The method of claim 38, further comprising dividing one or more of the geographic chunks into a plurality of tiles, the dynamic recognition grammar further organizing the grammar information according to the plurality of tiles.

44. The method of claim 43, further comprising dividing one or more of the plurality of tiles into a plurality of subtiles, the dynamic recognition grammar further organizing the grammar information according to the plurality of subtiles.

45. The method of claim 37, the dynamic recognition grammar having a size constrained according to memory or resource availability in the voice navigation services environment.

46. The method of claim 37, further comprising modifying the grammar information included in the dynamic recognition grammar in response to changes in the topological domains associated with the current location of the user.

47. The method of claim 37, further comprising determining affinities between the user and one or more peers of the user, the dynamic recognition grammar further organizing the grammar information according to the determined affinities.

48. A method for providing advertisements in a voice navigation services environment, comprising:

Google Exhibit 1019
Page 00718

identifying a current location of a user using a location detection system;

retrieving shared knowledge and information associated with a navigation context, the retrieved information used to determine probable interpretations of natural language utterances received in the voice navigation services environment;

identifying one or more advertisements to provide to the user, the one or more advertisements identified based on an affinity to one or more of the current location of the user, the shared knowledge, or the information associated with the navigation context; and

generating a multi-modal output to provide the one or more identified advertisements to the user.

49. The method of claim 48, the multi-modal output including a system-generated speech utterance.

50. The method of claim 48, the location detection system receiving information relating to the one or more advertisements over a data channel.

51. The method of claim 50, the data channel associated with a radio frequency identifier detected by the location detection system.

52. The method of claim 48, the shared knowledge including dialogue history information, request history information, user interface state information, short-term user profile information, long-term user profile information, peer user profile information, and user location information.

53. The method of claim 48, further comprising:

receiving a voice-based input that includes at least a natural language utterance;

using the shared knowledge and the information associated with the navigation context to determine a probable interpretation of the received natural language utterance, the advertisements identified based on the probable interpretation of the natural language utterance.

54. The method of claim 48, the advertisement including a guide to a community local to the current location of the user.

Google Exhibit 1019
Page 00719

55. The method of claim 54, the local community guide including recommendations for one or more points of interest, events, restaurants, stores, activities, or tourist attractions.

56. The method of claim 54, the local community guide including information associated with one or more of maps, destinations, directories, traffic, parking, or weather.

57. The method of claim 54, the local community guide based on affinities between the user and one or more of the current location of the user, one or more peers of the user, or one or more communities associated with the user.

58. The method of claim 48, the advertisement including one or more of an image, a banner, an audio message, a video message, a promotional offer, a coupon, and a data stream.

Google Exhibit 1019
Page 00720

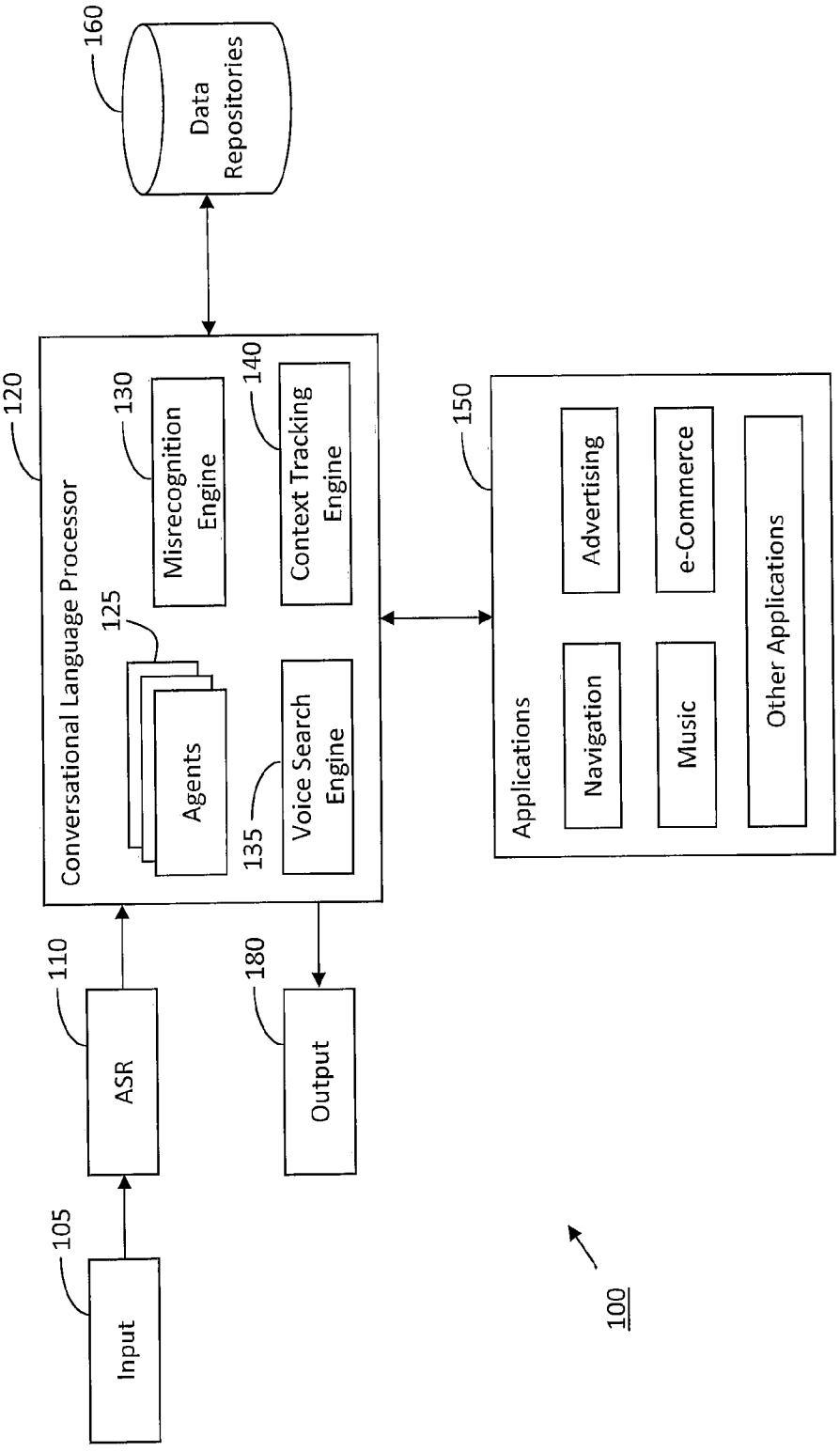

160
Data Repositories
120
Conversational Language Processor
130
Misrecognition Engine
140
Context Tracking Engine
125
Agents
135
Voice Search Engine
150
Applications
Advertising
e-Commerce
Navigation
Music
Other Applications
110
ASR
180
Output
105
Input
100

Figure 1

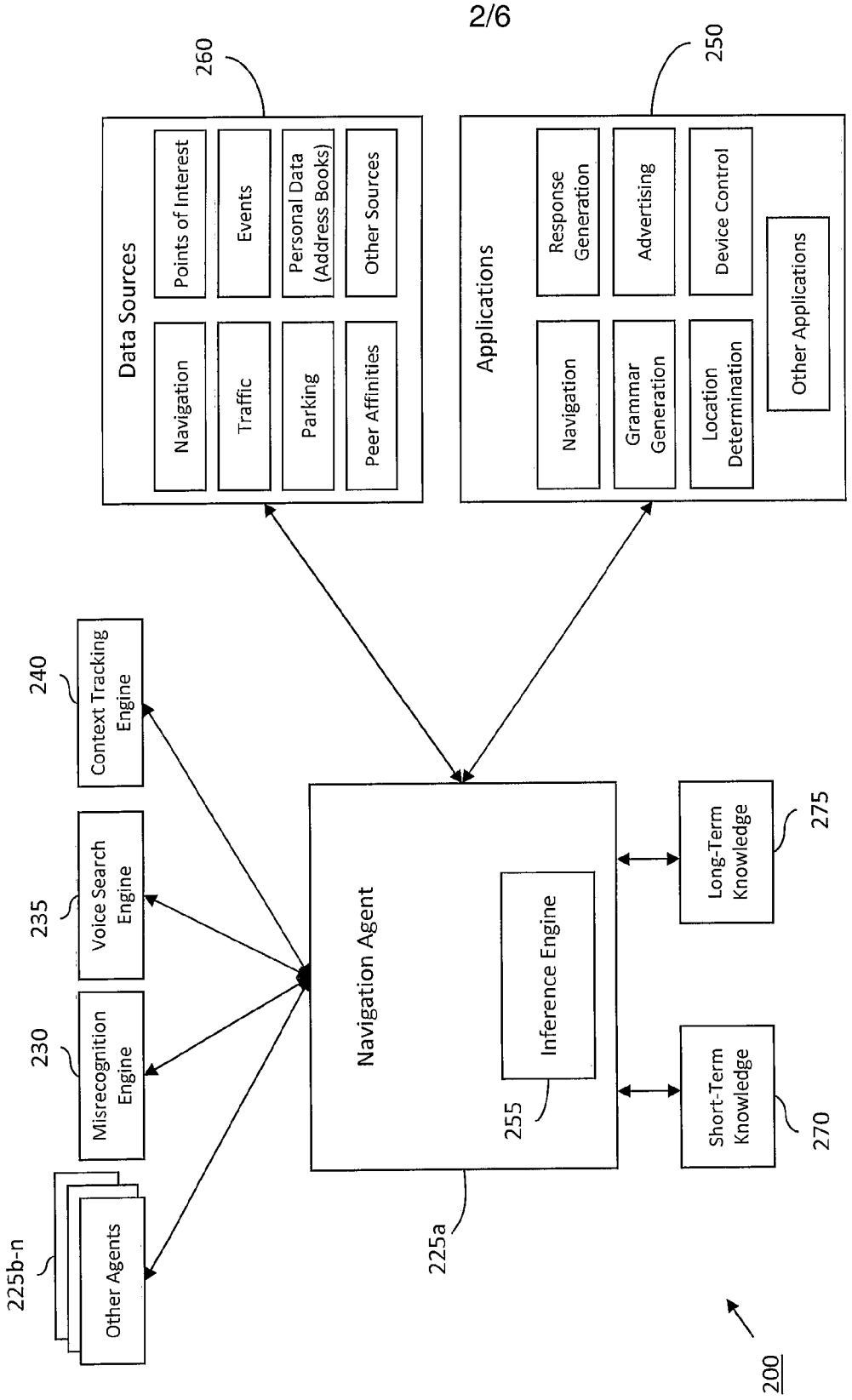

260
Data Sources
Points of Interest
Events
Personal Data (Address Books)
Other Sources
Navigation
Traffic
Parking
Peer Affinities
250
Applications
Response Generation
Advertising
Device Control
Navigation
Grammar Generation
Location Determination
Other Applications
240
Context Tracking Engine
235
Voice Search Engine
230
Misrecognition Engine
225b-n
Other Agents
Navigation Agent
Inference Engine
255
225a
Long-Term Knowledge
275
Short-Term Knowledge
270
200

Figure 2

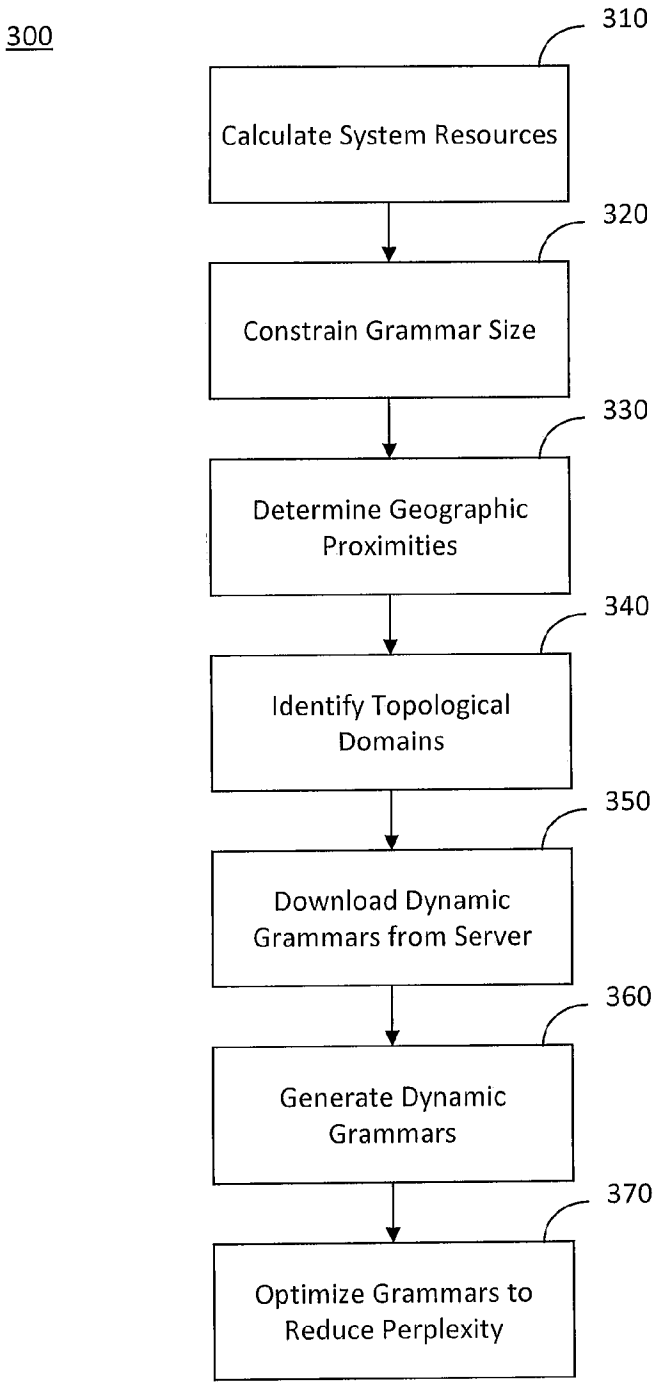

300
310
Calculate System Resources
320
Constrain Grammar Size
330
Determine Geographic Proximities
340
Identify Topological Domains
350
Download Dynamic Grammars from Server
360
Generate Dynamic Grammars
370
Optimize Grammars to Reduce Perplexity

Figure 3

400

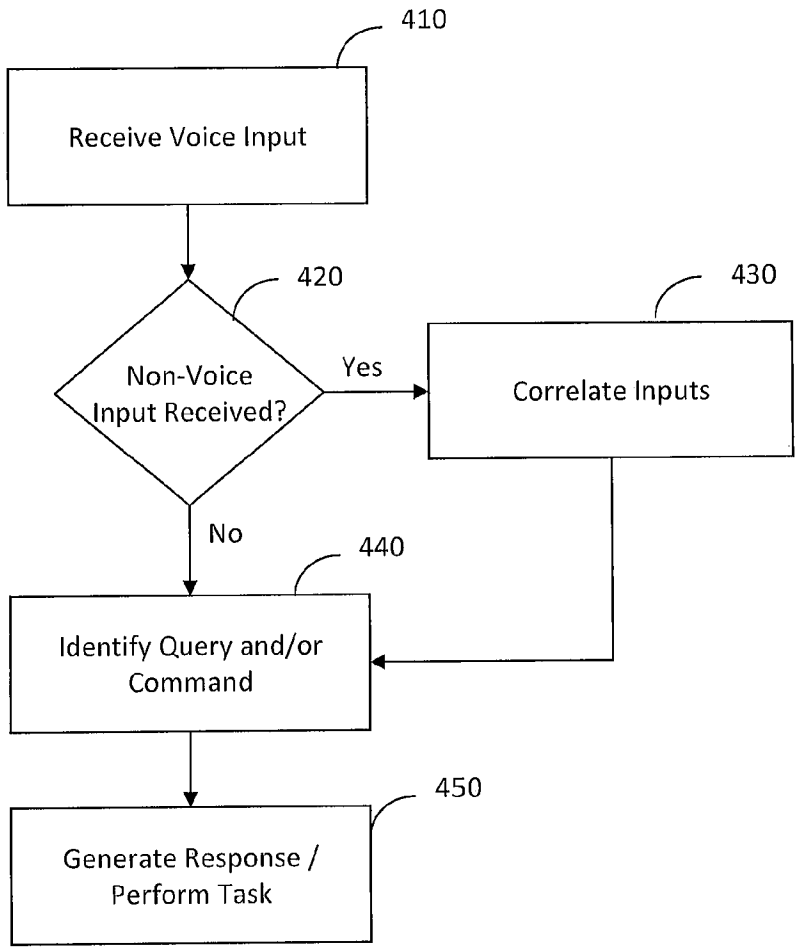

410
Receive Voice Input
420
Non-Voice Input Received?
Yes
430
Correlate Inputs
No
440
Identify Query and/or Command
450
Generate Response / Perform Task

Figure 4

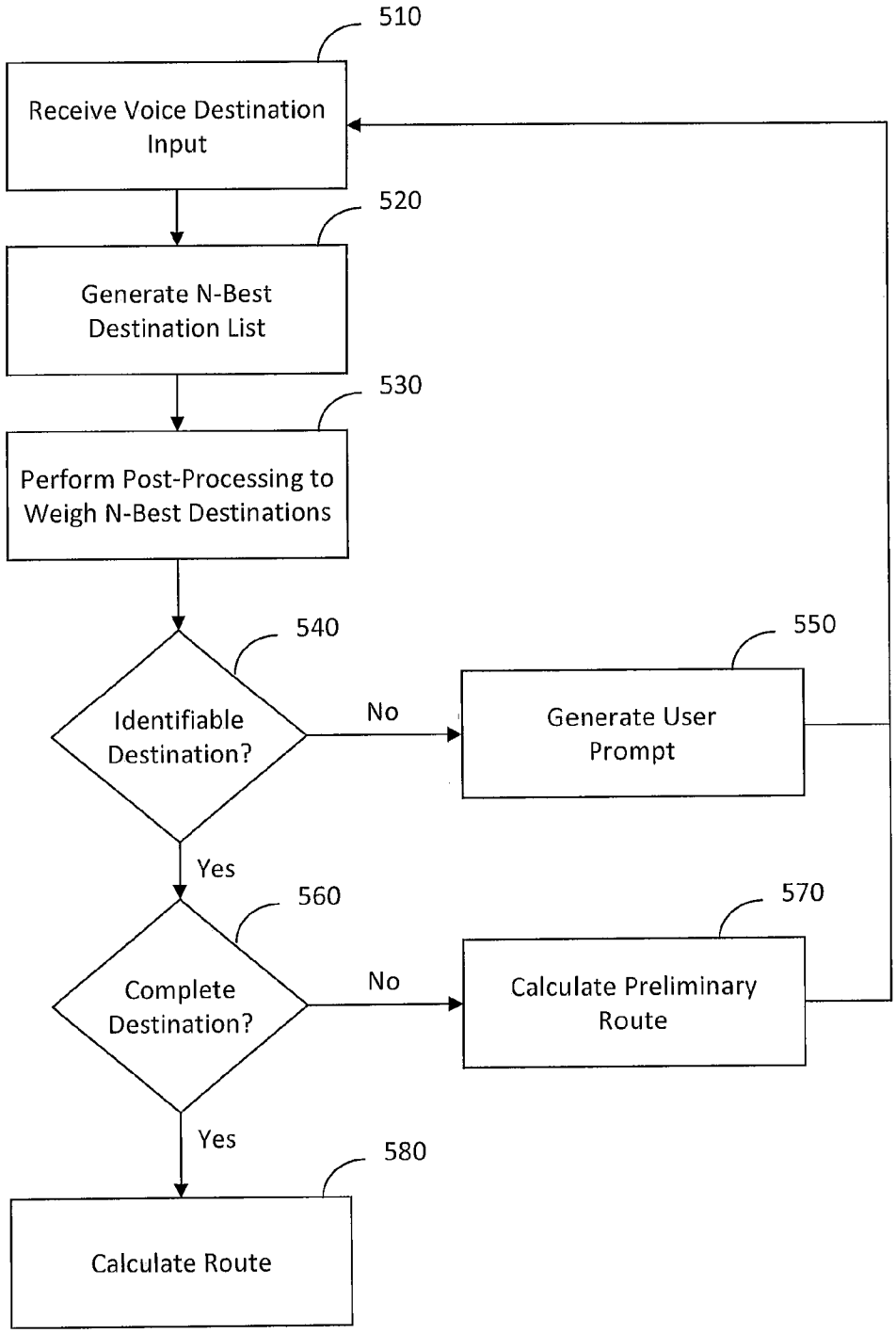

510
Receive Voice Destination Input
520
Generate N-Best Destination List
530
Perform Post-Processing to Weigh N-Best Destinations
540
Identifiable Destination?
No
550
Generate User Prompt
Yes
560
Complete Destination?
No
570
Calculate Preliminary Route
Yes
580
Calculate Route

Figure 5

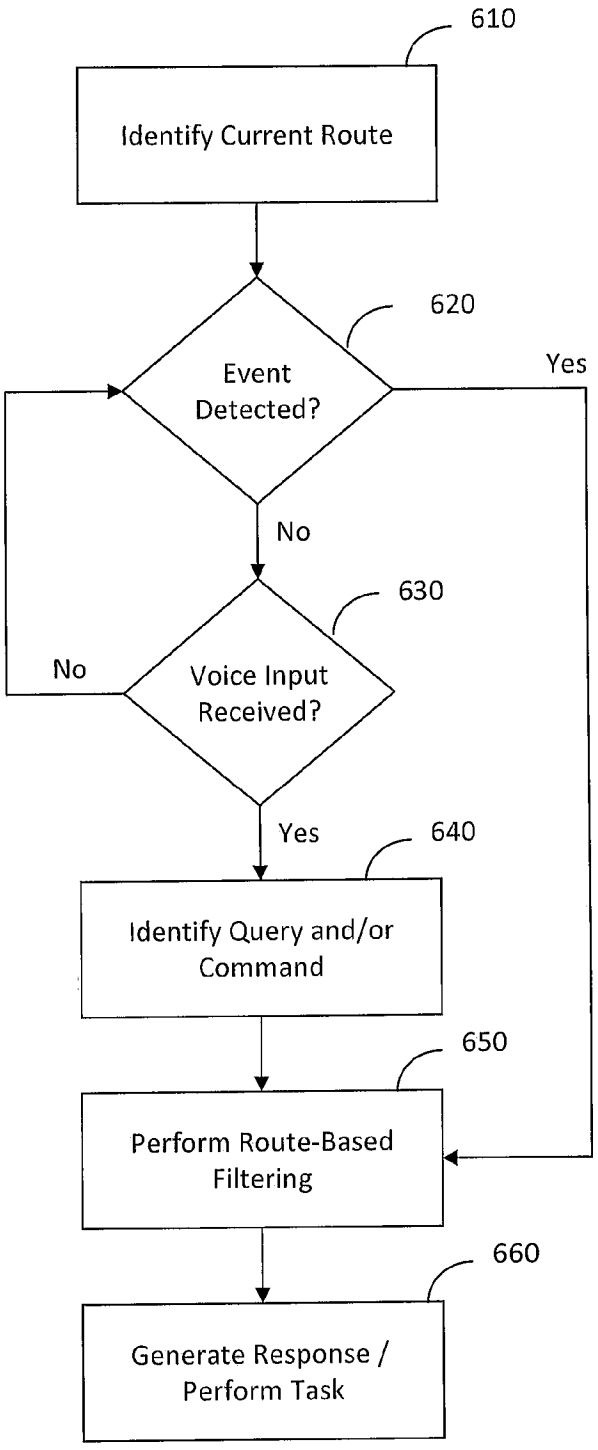

610
Identify Current Route
620
Event Detected?
Yes
No
630
Voice Input Received?
No
Yes
640
Identify Query and/or Command
650
Perform Route-Based Filtering
660
Generate Response / Perform Task

Figure 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 08/69508

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 17/27 (2008.04)
USPC - 704/9
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC - 704/9

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 704/1, 2, 8, 9, 200, 206, 221, E15.045

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (PGPB, USPT, EPAB, JPAB); Google Scholar
Search Terms Used: map, gps, speech, voice, talk, navigation, natural, conversational, grammar, interpret, software, computer, input, interface, user interface, refine, zoom, route, directions

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2007/0050191 A1 (WEIDER et al.) 01 March 2007 (01.03.2007), entire document, especially paras. [0061]-[0073], [0100], [0121], [0123], [0126], [0133], [0140], [0142], [0147], [0153], [0167], [0189]-[0194], [0210]-[0211], [0215]	1-42, 45-58 ---------- 43-44
Y	US 5,902,347 A (BACKMAN et al.) 11 May 1999 (11.05.1999), entire document, especially col 1 ln 32-34, col 4 ln 33-44	43-44
A	US 5,274,560 A (LARUE) 28 December 1993 (28.12.1993), entire document	1-58
A	US 6,598,018 B1 (JUNQUA) 22 July 2003 (22.07.2003), entire document	1-58

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 September 2008 (25.09.2008)	02 OCT 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 December 2009 (03.12.2009)

PCT

(10) International Publication Number
WO 2009/145796 A1

(51) International Patent Classification:
G06F 13/12 (2006.01)

(21) International Application Number:
PCT/US2008/069524

(22) International Filing Date:
9 July 2008 (09.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/127,343 27 May 2008 (27.05.2008) US

(71) Applicant *(for all designated States except US)*: **VOICEBOX TECHNOLOGIES, INC.** [US/US]; 11980 NE 24th Street, Bellevue, Washington 98005 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KENNEWICK, Robert** [US/US]; 13660 NE 42nd St., Bellevue, WA 98005 (US). **WEIDER, Chris** [US/US]; 10757 Greenwood Avenue N. #220, Seattle, Washington 98133 (US).

(74) Agents: **GATTO, James, G.** et al.; Pillsbury Winthrop Shaw Pittman LLP, P.O. Box 10500, McLean, Virginia 22102 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: SYSTEM AND METHOD FOR AN INTEGRATED, MULTI-MODAL, MULTI-DEVICE NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT

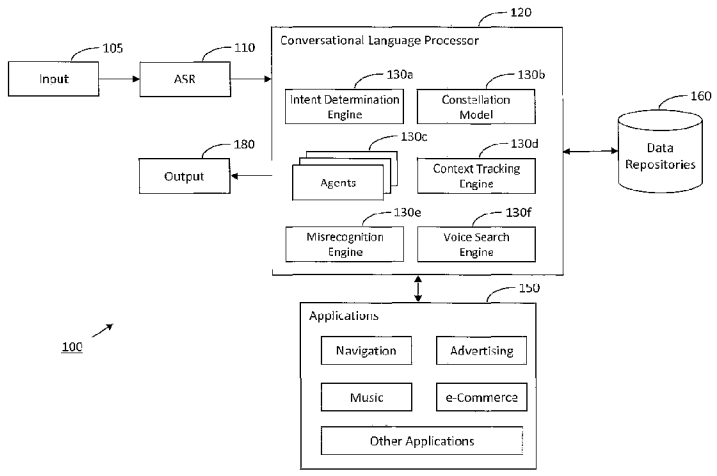


Figure 1

(57) **Abstract**: A system and method for an integrated, multi-modal, multi-device natural language voice services environment may be provided. In particular, the environment may include a plurality of voice-enabled devices each having intent determination capabilities for processing multi-modal natural language inputs in addition to knowledge of the intent determination capabilities of other devices in the environment. Further, the environment may be arranged in a centralized manner, a distributed peer-to-peer manner, or various combinations thereof. As such, the various devices may cooperate to determine intent of multi-modal natural language inputs, and commands, queries, or other requests may be routed to one or more of the devices best suited to take action in response thereto.

WO 2009/145796 A1

SYSTEM AND METHOD FOR AN INTEGRATED, MULTI-MODAL, MULTI-DEVICE NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT

FIELD OF THE INVENTION

[001] The invention relates to an integrated voice services environment in which a plurality of devices can provide various voice services by cooperatively processing free form, multi-modal, natural language inputs, thereby facilitating conversational interactions between a user and one or more of the devices in the integrated environment.

BACKGROUND OF THE INVENTION

[002] As technology has progressed in recent years, consumer electronic devices have emerged to become nearly ubiquitous in the everyday lives of many people. To meet the increasing demand that has resulted from growth in the functionality and mobility of mobile phones, navigation devices, embedded devices, and other such devices, a wealth of features and functions are often provided therein in addition to core applications. Greater functionality also introduces the trade-offs, however, including learning curves that often inhibit users from fully exploiting all of the capabilities of their electronic devices. For example, many existing electronic devices include complex human to machine interfaces that may not be particularly user-friendly, which inhibits mass-market adoption for many technologies. Moreover, cumbersome interfaces often result in otherwise desirable features being buried (e.g., within menus that may be tedious to navigate), which has the tendency of causing many users to not use, or even know about, the potential capabilities of their devices.

[003] As such, the increased functionality provided by many electronic devices often tends to be wasted, as market research suggests that many users only use only a fraction of the features or applications available on a given device. Moreover, in a society where wireless networking and broadband access are increasingly prevalent, consumers tend to naturally desire seamless mobile capabilities from their electronic devices. Thus, as consumer demand intensifies for simpler mechanisms to interact with electronic devices, cumbersome interfaces that prevent quick and focused interaction can become an important concern. Accordingly, the ever-growing demand for mechanisms to use technology in intuitive ways remains largely unfulfilled.

[004] One approach towards simplifying human to machine interactions in electronic devices includes the use of voice recognition software, which can enable users to exploit features that

Google Exhibit 1019
Page 00729

could otherwise be unfamiliar, unknown, or difficult to use. For example, a recent survey conducted by the Navteq Corporation, which provides data used in a variety of applications such as automotive navigation and web-based applications, demonstrates that voice recognition often ranks among the features most desired by consumers of electronic devices. Even so, existing voice user interfaces, when they actually work, still tend to require significant learning on the part of the user.

[005] For example, many existing voice user interface only support requests formulated according to specific command-and-control sequences or syntaxes. Furthermore, many existing voice user interfaces cause user frustration or dissatisfaction because of inaccurate speech recognition. Similarly, by forcing a user to provide pre-established commands or keywords to communicate requests in ways that a system can understand, existing voice user interfaces do not effectively engage the user in a productive, cooperative dialogue to resolve requests and advance a conversation towards a mutually satisfactory goal (e.g., when users may be uncertain of particular needs, available information, or device capabilities, among other things). As such, existing voice user interfaces tend to suffer from various drawbacks, including significant limitations on engaging users in a dialogue in a cooperative and conversational manner.

[006] Additionally, many existing voice user interfaces fall short in utilizing information distributed across various different domains or devices in order to resolve natural language voice-based inputs. Thus, existing voice user interfaces suffer from being constrained to a finite set of applications for which they have been designed, or to devices on which they reside. Although technological advancement has resulted in users often having several devices to suit their various needs, existing voice user interfaces do not adequately free users from device constraints. For example, users may be interested in services associated with different applications and devices, but existing voice user interfaces tend to restrict users from accessing the applications and devices as they see fit. Moreover, users typically can only practicably carry a finite number of devices at any given time, yet content or services associated with users' other devices that currently being used may be desired in various circumstances. Accordingly, although users tend to have varying needs, where content or services associated with different devices may be desired in various contexts or environments, existing voice technologies tend to fall short in providing an integrated environment in which users can request content or services associated with virtually any device or network. As such, constraints on information availability

Google Exhibit 1019
Page 00730

and device interaction mechanisms in existing voice services environments tend to prevent users from experiencing technology in an intuitive, natural, and efficient way.

[007] Existing systems suffer from these and other problems.

SUMMARY OF THE INVENTION

[008] According to various aspects of the invention, a system and method for an integrated, multi-modal, multi-device natural language voice services environment may include a plurality of voice-enabled devices each having intent determination capabilities for processing multi-modal natural language inputs in addition to knowledge of the intent determination capabilities of other devices in the environment. Further, the environment may be arranged in a centralized manner, a distributed peer-to-peer manner, or various combinations thereof. As such, the various devices may cooperate to determine intent of multi-modal natural language inputs, and commands, queries, or other requests may be routed to one or more of the devices best suited to take action in response thereto.

[009] According to various aspects of the invention, the integrated natural language voice services environment arranged in the centralized manner includes an input device that receives a multi-modal natural language input, a central device communicatively coupled to the input device, and one or more secondary devices communicatively coupled to the central device. Each of the input device, the central device, and the one or more secondary devices may have intent determination capabilities for processing multi-modal natural language inputs. As such, an intent of a given multi-modal natural language input may be determined in the centralized manner by communicating the multi-modal natural language input from the input device to the central device. Thereafter, the central device may aggregate the intent determination capabilities of the input device and the one or more secondary devices and determine an intent of the multi-modal natural language input using the aggregated intent determination capabilities. The input device may then receive the determined intent from the central device and invoke at least one action at one or more of the input device, the central device, or the secondary devices based on the determined intent.

[010] According to various aspects of the invention, the integrated natural language voice services environment arranged in the distributed manner includes an input device that receives a multi-modal natural language input, a central device communicatively coupled to the input

Google Exhibit 1019
Page 00731

device and one or more secondary devices communicatively coupled to the input device, wherein each of the input device and the one or more secondary devices may have intent determination capabilities for processing multi-modal natural language inputs, as in the centralized implementation. However, the distributed implementation may be distinct from the centralized implementation in that a preliminary intent of the multi-modal natural language input may be determined at the input device using local intent determination capabilities. The multi-modal natural language input may then be communicated to one or more of the secondary devices (e.g., when a confidence level of the intent determination at the input device falls below a given threshold). In such cases, each of the secondary devices determine an intent of the multi-modal natural language input using local intent determination capabilities. The input device collates the preliminary intent determination and the intent determinations of the secondary devices, and may arbitrate among the collated intent determinations to determine an actionable intent of the multi-modal natural input.

[011] According to various aspects of the invention, the integrated natural language voice services environment arranged in a manner that dynamically selects between a centralized model and a distributed model. For example, the environment includes an input device that receives a multi-modal natural language input one or more secondary devices communicatively coupled to the input device, each of which have intent determination capabilities for processing multi-modal natural language inputs. A constellation model may be accessible to each of the input device and the one or more secondary devices, wherein the constellation model describes the intent determination capabilities of the input device and the one or more secondary devices. The multi-modal natural language input can be routed for processing at one or more of the input device or the secondary devices to determine an intent thereof based on the intent determination capabilities described in the constellation model. For example, when the constellation model arranges the input device and the secondary devices in the centralized manner, one of the secondary devices may be designated the central device and the natural language input may be processed as described above. However, when the multi-modal natural language cannot be communicated to the central device, the constellation model may be dynamically rearranged in the distributed manner, whereby the input device and the secondary devices share knowledge relating to respective local intent determination capabilities and operate as cooperative nodes to

Google Exhibit 1019
Page 00732

determine the intent of the multi-modal natural language input using the shared knowledge relating to local intent determination capabilities.

[012] Other objects and advantages of the invention will be apparent based on the following drawings and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[013] Figure 1 illustrates a block diagram of an exemplary multi-modal electronic device that may be provided in an integrated, multi-device natural language voice services environment, according to various aspects of the invention.

[014] Figure 2 illustrates a block diagram of an exemplary centralized implementation of the integrated, multi-modal, multi-device natural language voice service environment, according to various aspects of the invention.

[015] Figure 3 illustrates a flow diagram of an exemplary method for processing multi-modal, natural language inputs at an input device in the centralized implementation of the integrated, multi-modal, multi-device natural language voice service environment, according to various aspects of the invention.

[016] Figure 4 illustrates a flow diagram of an exemplary method for processing multi-modal, natural language inputs at a central device in the centralized implementation of the integrated, multi-modal, multi-device natural language voice service environment, according to various aspects of the invention.

[017] Figure 5 illustrates a flow diagram of an exemplary method for processing multi-modal, natural language inputs at a secondary device in the centralized implementation of the integrated, multi-modal, multi-device natural language voice service environment, according to various aspects of the invention.

[018] Figure 6 illustrates a block diagram of an exemplary distributed implementation of the integrated, multi-modal, multi-device natural language voice service environment, according to various aspects of the invention.

[019] Figure 7 illustrates a flow diagram of an exemplary method for processing multi-modal, natural language inputs at an input device in the distributed implementation of the integrated, multi-modal, multi-device natural language voice service environment, according to various aspects of the invention.

Google Exhibit 1019
Page 00733

DETAILED DESCRIPTION

[020] According to various aspects of the invention, Figure 1 illustrates a block diagram of an exemplary multi-modal electronic device 100 that may be provided in a natural language voice services environment that includes one or more additional multi-modal devices (e.g., as illustrated in Figures 2 and 6). As will be apparent, the electronic device 100 illustrated in Figure 1 may be any suitable voice-enabled electronic device (e.g., a telematics device, a personal navigation device, a mobile phone, a VoIP node, a personal computer, a media device, an embedded device, a server, or another electronic device). The device 100 may include various components that collectively provide a capability to process conversational, multi-modal natural language inputs. As such, a user of the device 100 may engage in multi-modal conversational dialogues with the voice-enabled electronic device 100 to resolve requests in a free form, cooperative manner.

[021] For example, the natural language processing components may support free form natural language utterances to liberate the user from restrictions relating to how commands, queries, or other requests should be formulated. Rather, the user may employ any manner of speaking that feels natural in order to request content or services available through the device 100 (e.g., content or services relating to telematics, communications, media, messaging, navigation, marketing, information retrieval, etc.). For instance, in various implementations, the device 100 may process natural language utterances utilizing techniques described in co-pending U.S. Patent Application 10/452,147, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed June 3, 2003, and U.S. Patent Application 10/618,633, entitled "Mobile Systems and Methods for Responding to Natural Language Speech Utterance," filed June 15, 2003, the disclosures of which are hereby incorporated by reference in their entirety.

[022] Moreover, because the device 100 may be deployed in an integrated multi-device environment, the user may further request content or services available through other devices deployed in the environment. In particular, the integrated voice services environment may include a plurality of multi-modal devices, each of which include natural language components generally similar to those illustrated in Figure 1. The various devices in the environment may serve distinct purposes, however, such that available content, services, applications, or other capabilities may vary among the devices in the environment (e.g., core functions of a media

Google Exhibit 1019
Page 00734

device may vary from those of a personal navigation device). Thus, each device in the environment, including device 100, may have knowledge of content, services, applications, intent determination capabilities, and other features available through the other devices by way of a constellation model 130b. Accordingly, as will be described in greater detail below, the electronic device 100 may cooperate with other devices in the integrated environment to resolve requests by sharing context, prior information, domain knowledge, short-term knowledge, long-term knowledge, and cognitive models, among other things.

[023] According to various aspects of the invention, the electronic device 100 may include an input mechanism 105 that can receive multi-modal natural language inputs, which include at least an utterance spoken by the user. As will be apparent, the input mechanism 105 may include any appropriate device or combination of devices capable of receiving a spoken input (e.g., a directional microphone, an array of microphones, or any other device that can generate encoded speech). Further, in various implementations, the input mechanism 105 can be configured to maximize fidelity of encoded speech, for example, by maximizing gain in a direction of the user, cancelling echoes, nulling point noise sources, performing variable rate sampling, or filtering environmental noise (e.g., background conversations). As such, the input mechanism 105 may generate encoded speech in a manner that can tolerate noise or other factors that could otherwise interfere with accurate interpretation of the utterance.

[024] Furthermore, in various implementations, the input mechanism 105 may include various other input modalities (i.e., the input mechanism 105 may be arranged in a multi-modal environment), in that non-voice inputs can be correlated and/or processed in connection with one or more previous, contemporaneous, or subsequent multi-modal natural language inputs. For example, the input mechanism 105 may be coupled to a touch-screen interface, a stylus and tablet interface, a keypad or keyboard, or any other suitable input mechanism, as will be apparent. As a result, an amount of information potentially available when processing the multi-modal inputs may be maximized, as the user can clarify utterances or otherwise provide additional information in a given multi-modal natural language input using various input modalities. For instance, in an exemplary illustration, the user could touch a stylus or other pointing device to a portion of a touch-screen interface of the device 100, while also providing an utterance relating to the touched portion of the interface (e.g., "Show me restaurants around here"). In this example, the natural language utterance may be correlated with the input received

Google Exhibit 1019
Page 00735

via the touch-screen interface, resulting in "around here" being interpreted in relation to the touched portion of the interface (e.g., as opposed to the user's current location or some other meaning).

[025] According to various aspects of the invention, the device 100 may include an Automatic Speech Recognizer 110 that generates one or more preliminary interpretations of the encoded speech, which may be received from the input mechanism 105. For example, the Automatic Speech Recognizer 110 may recognize syllables, words, or phrases contained in an utterance using one or more dynamically adaptable recognition grammars. The dynamic recognition grammars may be used to recognize a stream of phonemes through phonetic dictation based on one or more acoustic models. Furthermore, as described in co-pending U.S. Patent Application 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed August 5, 2005, the disclosure of which is hereby incorporated by reference in its entirety, the Automatic Speech Recognizer 110 may be capable of multi-pass analysis, where a primary speech recognition engine may generate a primary interpretation of an utterance (e.g., using a large list dictation grammar) and request secondary transcription from one or more secondary speech recognition engines (e.g., using a virtual dictation grammar having decoy words for out-of-vocabulary words).

[026] Thus, the Automatic Speech Recognizer 110 may generate preliminary interpretations of an utterance in various ways, including exclusive use of a dictation grammar or virtual dictation grammar, or use of various combinations thereof (e.g., when the device 100 supports multi-pass analysis). In any event, the Automatic Speech Recognizer 110 may provide out-of-vocabulary capabilities and may tolerate portions of a speech signal being dropped, the user misspeaking, or other variables that may occur in natural language speech (e.g., stops and starts, stutters, etc.). Furthermore, the recognition grammars employed by the Automatic Speech Recognizer 110 may include vocabularies, dictionaries, syllables, words, phrases, or other information optimized according to various contextual or application-specific domains (e.g., navigation, music, movies, weather, shopping, news, languages, temporal or geographic proximities, or other suitable domains). Moreover, environmental knowledge (e.g., peer-to-peer affinities, capabilities of devices in the environment, etc.), historical knowledge (e.g., frequent requests, prior context, etc.), or other types of knowledge can be used to continually optimize the information contained in the recognition grammars on a dynamic basis.

Google Exhibit 1019
Page 00736

[027] For example, information contained in the recognition grammars may be dynamically optimized to improve a likelihood of a given utterance being recognized accurately (e.g., following an incorrect interpretation of a word, the incorrect interpretation may be removed from the grammar to reduce a likelihood of the incorrect interpretation being repeated). Accordingly, the Automatic Speech Recognizer 110 may use a number of techniques to generate preliminary interpretations of natural language utterances, including those described, for example, in co-pending U.S. Patent Application 11/513,269, entitled "Dynamic Speech Sharpening," filed August 31, 2006, the disclosure of which is hereby incorporated by reference in its entirety. Furthermore, the techniques used by the Automatic Speech Recognizer 110 associated with the device 100 may be considered in defining intent determination capabilities of the device 100, and such capabilities may be shared with other devices in the environment to enable convergence of speech recognition throughout the environment (e.g., because various devices may employ distinct speech recognition techniques or have distinct grammars or vocabularies, the devices may share vocabulary translation mechanisms to enhance system-wide recognition).

[028] According to various aspects of the invention, the Automatic Speech Recognizer 110 may provide one or more preliminary interpretations of a multi-modal input, including an utterance contained therein, to a conversational language processor 120. The conversational language processor 120 may include various components that collectively operate to model everyday human-to-human conversations in order to engage in cooperative conversations with the user to resolve requests based on the user's intent. For example, the conversational language processor 120 may include, among other things, an intent determination engine 130a, a constellation model 130b, one or more domain agents 130c, a context tracking engine 130d, a misrecognition engine 130e, and a voice search engine 130f. Furthermore, the conversational language processor 120 may be coupled to one or more data repositories 160 and applications associated with one or more domains. Thus, the intent determination capabilities of the device 100 may be defined based on the data and processing capabilities of the Automatic Speech Recognizer 110 and the conversational language processor 120.

[029] More particularly, the intent determination engine 130a may establish meaning for a given multi-modal natural language input based on a consideration of the intent determination capabilities of the device 100 as well as the intent determination capabilities of other devices in the integrated voice services environment. For example, the intent determination capabilities of

Google Exhibit 1019
Page 00737

the device 100 may be defined as a function of processing resources, storage for grammars, context, agents, or other data, and content or services associated with the device 100 (e.g., a media device with a small amount of memory may have a smaller list of recognizable songs than a device with a large amount of memory). Thus, the intent determination engine 130a may determine whether to process a given input locally (e.g., when the device 100 has intent determination capabilities that suggest favorable conditions for recognition), or whether to route information associated with the input to other devices, which may assist in determining the intent of the input.

[030] As such, to determine which device or combination of devices should process an input, the intent determination engine 130a may evaluate the constellation model 130b, which provides a model of the intent determination capabilities for each of the devices in the integrated voice services environment. For instance, the constellation model 130b may contain, among other things, knowledge of processing and storage resources available to each of the devices in the environment, as well as the nature and scope of domain agents, context, content, services, and other information available to each of the devices in the environment. As such, using the constellation model 130b, the intent determination engine 130a may be able to determine whether any of the other devices have intent determination capabilities that can be invoked to augment or otherwise enhance the intent determination capabilities of the device 100 (e.g., by routing information associated with a multi-modal natural language input to the device or devices that appear best suited to analyze the information and therefore determine an intent of the input). Accordingly, the intent determination engine 130a may establish the meaning of a given utterance by utilizing the comprehensive constellation model 130b that describes capabilities within the device 100 and across the integrated environment. The intent determination engine 130a may therefore optimize processing of a given natural language input based on capabilities throughout the environment (e.g., utterances may be processed locally to the device 100, routed to a specific device based on information in the constellation model 130b, or flooded to all of the devices in the environment in which case an arbitration may occur to select a best guess at an intent determination).

[031] Although the following discussion will generally focus on various techniques that can be used to determine the intent of multi-modal natural language inputs in the integrated multi-device environment, it will be apparent that the natural language processing capabilities of any

Google Exhibit 1019
Page 00738

one of the devices may extend beyond the specific discussion that has been provided herein. As such, in addition to the co-pending U.S. Patent Applications referenced above, further natural language processing capabilities that may be employed include those described in co-pending U.S. Patent Application 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed August 5, 2005, U.S. Patent Application 11/200,164, entitled "System and Method of Supporting Adaptive Misrecognition in Conversational Speech," filed August 10, 2005, U.S. Patent Application 11/212,693, entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," filed August 29, 2005, U.S. Patent Application 11/580,926, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed October 16, 2006, U.S. Patent Application 11/671,526, entitled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input," filed February 6, 2007, and U.S. Patent Application 11/954,064, entitled "System and Method for Providing a Natural Language Voice User Interface in an Integrated Voice Navigation Services Environment," filed December 11, 2007, the disclosures of which are hereby incorporated by reference in their entirety.

[032] According to various aspects of the invention, Figure 2 illustrates a block diagram of an exemplary centralized implementation of the integrated, multi-modal, multi-device natural language voice service environment. As will be apparent from the further description to be provided herein, the centralized implementation of the integrated, multi-device voice services environment may enable a user to engage in conversational, multi-modal natural language interactions with any one of voice-enabled devices 210a-n or central voice-enabled device 220. As such, the multi-device voice services environment may collectively determine intent for any given multi-modal natural language input, whereby the user may request content or voice services relating to any device or application in the environment, without restraint.

[033] As illustrated in Figure 2, the centralized implementation of the multi-device voice service environment may include a plurality of voice-enabled devices 210a-n, each of which include various components capable of determining intent of natural language utterances, as described above in reference to Figure 1. Furthermore, as will be apparent, the centralized implementation includes a central device 220, which contains information relating to intent determination capabilities for each of the other voice-enabled devices 210a-n. For example, in various exemplary implementations, the central device 220 may be designated as such by virtue

Google Exhibit 1019
Page 00739

of being a device most capable of determining the intent of an utterance (e.g., a server, home data center, or other device having significant processing power, memory resources, and communication capabilities making the device suitable to manage intent determination across the environment). In another exemplary implementation, the central device 220 may be dynamically selected based on one or more characteristics of a given multi-modal natural language input, dialogue, or interaction (e.g., a device may be designated as the central device 220 when a current utterance relates to a specific domain).

[034] In the centralized implementation illustrated in Figure 2, a multi-modal natural language input may be received at one of the voice-enabled devices 210a-n. Therefore, the receiving one of the devices 210a-n may be designated as an input device for that input, while the remaining devices 210a-n may be designated as secondary devices for that input. In other words, for any given multi-modal natural language input, the multi-device environment may include an input device that collects the input, a central device 220 that aggregates intent determination, inferencing, and processing capabilities for all of the devices 210a-n in the environment, and one or more secondary devices that may also be used in the intent determination process. As such, each device 210 in the environment may be provided with a constellation model that identifies all of the devices 210 having incoming and outgoing communication capabilities, thus indicating an extent to which other devices may be capable of determining intent for a given multi-modal natural language input. The constellation model may further define a location of the central device 220, which aggregates context, vocabularies, content, recognition grammars, misrecognitions, shared knowledge, intent determination capabilities, inferencing capabilities, and other information from the various devices 210a-n in the environment.

[035] Accordingly, as communication and processing capabilities permit, the central device 220 may be used as a recognizer of first or last resort. For example, because the central device 220 converges intent determination capabilities across the environment (e.g., by aggregating context, vocabularies, device capabilities, and other information from the devices 210a-n in the environment), inputs may be automatically routed to the central device 220 when used as a recognizer of first resort, or as a recognizer of last resort when local processing at the input device 210 cannot determine the intent of the input with a satisfactory level of confidence. However, it will also be apparent that in certain instances the input device 210 may be unable to make contact with the central device 220 for various reasons (e.g., a network connection may be

Google Exhibit 1019
Page 00740

unavailable, or a processing bottleneck at the central device 220 may cause communication delays). In such cases, the input device 210 that has initiated contact with the central device 220 may shift into decentralized processing (e.g., as described in reference to Figure 6) and communicate capabilities with one or more of the other devices 210a-n in the constellation model. Thus, when the central device 220 cannot be invoked for various reasons, the remaining devices 210a-n may operate as cooperative nodes to determine intent in a decentralized manner.

[036] Additionally, in the multi-device voice services environment, the central device 220 and the various other devices 210a-n may cooperate to create a converged model of capabilities throughout the environment. For example, as indicated above, in addition to having intent determination capabilities based on processing resources, memory resources, and device capabilities, each of the devices 210a-n and the central device 220 may include various other natural language processing components. The voice services environment may therefore operate in an integrated manner by maintaining not only a complete model of data, content, and services associated with the various devices 210a-n, but also of other natural language processing capabilities and dynamic states associated with the various devices 210a-n. As such, the various devices 210a-n may operate with a goal of converging capabilities, data, states, and other information across the device, either on one device (e.g., the central device 220) or distributed among the various devices 210a-n (e.g., as in the decentralized implementation to be described in Figure 6).

[037] For example, as discussed above, each device 210 includes an Automatic Speech Recognizer, one or more dynamically adaptable recognition grammars, and vocabulary lists used to generate phonemic interpretations of natural language utterances. Moreover, each device 210 includes locally established context, which can range from information contained in a context stack, context and namespace variables, vocabulary translation mechanisms, short-term shared knowledge relating to a current dialogue or conversational interaction, long-term shared knowledge relating to a user's learned preferences over time, or other contextual information. Furthermore, each device 210 may have various services or applications associated therewith, and may perform various aspects of natural language processing locally. Thus, additional information to be converged throughout the environment may include partial or preliminary utterance recognitions, misrecognitions or ambiguous recognitions, inferencing capabilities, and

Google Exhibit 1019
Page 00741

overall device state information (e.g., songs playing in the environment, alarms set in the environment, etc.).

[038] Thus, various data synchronization and referential integrity algorithms may be employed in concert by the various devices 210a-n and the central device 220 to provide a consistent worldview of the environment. For example, information may be described and transmitted throughout the environment for synchronization and convergence purposes using the Universal Plug and Play protocol designed for computer ancillary devices, although the environment can also operate in a peer-to-peer disconnected mode (e.g., when the central device 220 cannot be reached). However, in various implementations, the environment may also operate in a peer-to-peer mode regardless of the disconnected status, as illustrated in Figure 6, for example, when the devices 210a-n have sufficient commensurate resources and capabilities for natural language processing.

[039] In general, the algorithms for convergence in the environment can be executed at various intervals, although it may be desirable to limit data transmission so as to avoid processing bottlenecks. For example, because the convergence and synchronization techniques relate to natural language processing, in which any given utterance will typically be expressed over a course of several seconds, information relating to context and vocabulary need not be updated on a time frame of less than a few seconds. However, as communication capabilities permit, context and vocabulary could be updated more frequently to provide real-time recognition or the appearance of real-time recognition. In another implementation, the convergence and synchronization may be permitted to run until completion (e.g., when no requests are currently pending), or the convergence and synchronization may be suspended or terminated when a predetermined time or resource consumption limit has been reached (e.g., when the convergence relates to a pending request, an intent determination having a highest confidence level at the time of cut-off may be used).

[040] By establishing a consistent view of capabilities, data, states, and other information throughout the environment, an input device 210 may cooperate with the central device 220 and one or more secondary devices (i.e., one or more of devices 210a-n, other than the input device) in processing any given multi-modal natural language input. Furthermore, by providing each device 210 and the central device 220 with a constellation model that describes a synchronized state of the environment, the environment may be tolerant of failure by one or more of the

Google Exhibit 1019
Page 00742

devices 210a-n, or of the central device 220. For example, if the input device 210 cannot communicate with the central device 220 (e.g., because of a server crash), the input device 210 may enter a disconnected peer-to-peer mode, whereby capabilities can be exchanged with one or more devices 210a-n with which communications remain available. To that end, when a device 210 establishes new information relating to vocabulary, context, misrecognitions, agent adaptation, intent determination capabilities, inferencing capabilities, or otherwise, the device 210 may transmit the information to the central device 220 for convergence purposes, as discussed above, in addition to consulting the constellation model to determine whether the information should be transmitted to one or more of the other devices 210a-n.

[041] For example, suppose the environment includes a voice-enabled mobile phone that has nominal functionality relating to playing music or other media, and which further has a limited amount of local storage space, while the environment further includes a voice-enabled home media system that includes a mass storage medium that provides dedicated media functionality. If the mobile phone were to establish new vocabulary, context, or other information relating to a song (e.g., a user downloads the song or a ringtone to the mobile phone while on the road), the mobile phone may transmit the newly established information to the home media system in addition to the central device 220. As such, by having a model of all of the devices 210a-n in the environment and transmitting new information to the devices where it will most likely be useful, the various devices may handle disconnected modes of operation when the central device 220 may be unavailable for any reason, while resources may be allocated efficiently throughout the environment.

[042] Thus, based on the foregoing discussion, it will be apparent that a centralized implementation of an integrated multi-device voice services environment may generally include a central device 220 operable to aggregate or converge knowledge relating to content, services, capabilities, and other information associated with various voice-enabled devices 210a-n deployed within the environment. In such centralized implementations, the central device 220 may be invoked as a recognizer of first or last resort, as will be described in greater detail with reference to Figures 3-5, and furthermore, the other devices 210a-n in the environment may be configured to automatically enter a disconnected or peer-to-peer mode of operation when the central device 220 cannot be invoked for any reason (i.e., devices may enter a decentralized or distributed mode, as will be described in greater detail with reference to Figures 6-7).

Google Exhibit 1019
Page 00743

Knowledge and capabilities of each of the devices 210a-n may therefore be made available throughout the voice services environment in a centralized manner, a distributed manner, or various combinations thereof, thus optimizing an amount of natural language processing resources used to determine an intent of any given multi-modal natural language input.

[043] According to various aspects of the invention, Figure 3 illustrates a flow diagram of an exemplary method for processing multi-modal, natural language inputs at an input device in the centralized implementation of the integrated, multi-modal, multi-device natural language voice service environment. Similarly, Figures 4 and 5 illustrate corresponding methods associated with a central device and one or more secondary devices, respectively, in the centralized voice service environment. Furthermore, it will be apparent that the processing techniques described in relation to Figures 3-5 may generally be based on the centralized implementation illustrated in Figure 2 and described above, whereby the input device may be assumed to be distinct from the central device, and the one or more secondary devices may be assumed to be distinct from the central device and the input device. However, it will be apparent that various instances may involve a natural language input being received at the central device or at another device, in which case the techniques described in Figures 3-5 may be vary depending on circumstances of the environment (e.g., decisions relating to routing utterances to a specific device or devices may be made locally, collaboratively, or in other ways depending on various factors, such as overall system state, communication capabilities, intent determination capabilities, or otherwise).

[044] As illustrated in Figure 3, a multi-modal natural language input may be received at an input device in an operation 310. The multi-modal input may include at least a natural language utterance provided by a user, and may further include other input modalities such as audio, text, button presses, gestures, or other non-voice inputs. It will also be apparent that prior to receiving the natural language input in operation 310, the input device may be configured to establish natural language processing capabilities. For example, establishing natural language processing capabilities may include, among other things, loading an Automatic Speech Recognizer and any associated recognition grammars, launching a conversational language processor to handle dialogues with the user, and installing one or more domain agents that provide functionality for respective application domains or contextual domains (e.g., navigation, music, movies, weather, information retrieval, device control, etc.).

Google Exhibit 1019
Page 00744

[045] The input device may also be configured to coordinate synchronization of intent determination capabilities, shared knowledge, and other information with the central device and the secondary devices in the environment prior to receiving the input at operation 310. For example, when the input device installs a domain agent, the installed domain agent may bootstrap context variables, semantics, namespace variables, criteria values, and other context related to that agent from other devices in the system. Similarly, misrecognitions may be received from the central device and the secondary devices in order to enable correction of agents that use information relevant to the received misrecognitions, and vocabularies and associated translation mechanisms may be synchronized among the devices to account for potential variations between the Automatic Speech Recognizers used by the various devices (e.g., each device in the environment cannot be guaranteed to use the same Automatic Speech Recognizer or recognition grammars, necessitating vocabulary and translation mechanisms to be shared among the devices that share intent determination capabilities).

[046] Upon establishing and synchronizing natural language processing capabilities and subsequently receiving a multi-modal natural language input in operation 310, the input device may determine whether the environment has been set up to automatically transmit the input to the central device in a decisional operation 320. In such a case, processing proceeds to an operation 360 for transmitting the input to the central device, which may then process the input according to techniques to be described in relation to Figure 4. If the environment has not been set up to automatically communicate the input to the central device, however, processing proceeds to an operation 330, where the input device performs transcription of the natural language utterance contained in the multi-modal input. For example, the input device may transcribe the utterance using the Automatic Speech Recognizer and recognition grammars associated therewith according to techniques described above and in the above-referenced U.S. Patent Applications.

[047] Subsequently, in an operation 340, an intent of the multi-modal natural language input may be determined at the input device using local natural language processing capabilities and resources. For example, any non-voice input modalities included in the input may be merged with the utterance transcription and a conversational language processor associated with the input device may utilize local information relating to context, domain knowledge, shared knowledge, context variables, criteria values, or other information useful in natural language

Google Exhibit 1019
Page 00745

processing. As such, the input device may attempt to determine a best guess as to an intent of the user that provided the input, such as identifying a conversation type (e.g., query, didactic, or exploratory) or request that may be contained in the input (e.g., a command or query relating to one or more domain agents or application domains).

[048] The intent determination of the input device may be assigned a confidence level (e.g., a device having an Automatic Speech Recognizer that implements multi-pass analysis may assign comparatively higher confidence levels to utterance transcriptions created thereby, which may result in a higher confidence level for the intent determination). The confidence level may be assigned based on various factors, as described in the above-referenced U.S. Patent Applications. As such, a decisional operation 350 may include determining whether the intent determination of the input device meets an acceptable level of confidence. When the intent determination meets the acceptable level confidence, processing may proceed directly to an operation 380 where action may be taken in response thereto. For example, when the intent determination indicates that the user has requested certain information, one or more queries may be formulated to retrieve the information from appropriate information sources, which may include one or more of the other devices. In another example, when the intent determination indicates that the user has requested a given command (e.g., to control a specific device), the command may be routed to the appropriate device for execution.

[049] Thus, in cases where the input device can determine the intent of a natural language input without assistance from the central device or the secondary devices, communications and processing resources may be conserved by taking immediate action as may be appropriate. On the other hand, when the intent determination of the input device does not meet the acceptable level of confidence, decisional operation 350 may result in the input device requesting assistance from the central device in operation 360. In such a case, the multi-modal natural language input may be communicated to the central device in its entirety, whereby the central device processes the input according to techniques described in Figure 4. However, should transmission to the central device fail for some reason, the input device may shift into a disconnected peer-to-peer mode where one or more secondary devices may be utilized, as will be described below in relation to Figure 7. When transmission to the central device occurs without incident, however, the input device may receive an intent determination from the central device in an operation 370, and may further receive results of one or more requests that the central device was able to

Google Exhibit 1019
Page 00746

resolve, or requests that the central device has formulated for further processing on the input device. As such, the input device may take action in operation 380 based on the information received from the central device in operation 370. For example, the input device may route queries or commands to local or remote information sources or devices based on the intent determination, or may present results of the requests processed by the central device to the user.

[050] Referring to Figure 4, the central device may receive the multi-modal natural language input from the input device in an operation 410. The central device, having aggregated context and other knowledge from throughout the environment, may thus transcribe the utterance in an operation 420 and determine an intent of the input from the transcribed utterance in an operation 430. As such, the central device may consider information relating to context, domain agents, applications, and device capabilities throughout the environment in determining the intent of the utterance, including identification of one or more domains relevant to the input. However, it will be apparent that utilizing information aggregated from throughout the environment may cause ambiguity or uncertainty in various instances (e.g., an utterance containing the word "traffic" may have a different intent in domains relating to movies, music, and navigation).

[051] As such, once the central device has attempted to determine the intent of the natural language input, a determination may be made in an operation 440 as to whether one or more secondary devices (i.e., other devices in the constellation besides the input device) may also be capable of intent determination in the identified domain or domains. When no such secondary devices can be identified, decisional operation 440 may branch directly to an operation 480 to send to the input device the determined intent and any commands, queries, or other requests identified from the determined intent.

[052] On the other hand, when one or more secondary devices in the environment have intent determination capabilities in the identified domain or domains, the natural language input may be sent to such secondary devices in an operation 450. The secondary devices may then determine an intent as illustrated in Figure 5, which may include techniques generally similar to those described above in relation to the input device and central device (i.e., the natural language input may be received in an operation 510, an utterance contained therein may be transcribed in an operation 520, and an intent determination made in an operation 530 may be returned to the central device in an operation 540).

Google Exhibit 1019
Page 00747

[053] Returning to Figure 4, the central device may collate intent determination responses received from the secondary devices in an operation 460. For example, as indicated above, the central device may identify one or more secondary devices capable of determining intent in a domain that the central device has identified as being relevant to the natural language utterance. As will be apparent, the secondary devices invoked in operation 450 may often include a plurality of devices, and intent determination responses may be received from the secondary devices in an interleaved manner, depending on processing resources, communications throughput, or other factors (e.g., the secondary devices may include a telematics device having a large amount of processing power and a broadband network connection and an embedded mobile phone having less processing power and only a cellular connection, in which case the telematics device may be highly likely to provide results to the central device before the embedded mobile phone). Thus, based on potential variations in response time of secondary devices, the central device may be configured to place constraints on collating operation 460. For example, the collating operation 460 may be terminated as soon as an intent determination has been received from one of the secondary devices that meets an acceptable level of confidence, or the operation 460 may be cut off when a predetermined amount of time has lapsed or a predetermined amount of resources have been consumed. In other implementations, however, it will be apparent that collating operation 460 may be configured to run to completion, regardless of whether delays have occurred or suitable intent determinations have been received. Further, it will be apparent that various criteria may be used to determine whether or when to end the collating operation 460, including the nature of a given natural language input, dialogue, or other interaction, or system or user preferences, among other criteria, as will be apparent.

[054] In any event, when the collating operation 460 has completed, a subsequent operation 470 may include the central device arbitrating among the intent determination responses received from one or more of the secondary devices previously invoked in operation 450. For example, each of the invoked secondary devices that generate an intent determination may also assign a confidence level to that intent determination, and the central device may consider the confidence levels in arbitrating among the responses. Moreover, the central device may associate other criteria with the secondary devices or the intent determinations received from the secondary devices to further enhance a likelihood that the best intent determination will be used. For example, various ones of the secondary devices may only be invoked for partial recognition in

Google Exhibit 1019
Page 00748

distinct domains, and the central device may aggregate and arbitrate the partial recognitions to create a complete transcription. In another example, a plurality of secondary devices may be invoked to perform overlapping intent determination, and the central device may consider capabilities of the secondary devices to weigh the respective confidence levels (e.g., when one of two otherwise identical secondary devices employs multi-pass speech recognition analysis, the secondary device employing the multi-pass speech recognition analysis may be weighed as having a higher likelihood of success). It will be apparent that the central device may be configured to arbitrate and select one intent determination from among all of the intent hypotheses, which may include the intent determination hypothesis generated by the central device in operation 430. Upon selecting the best intent determination hypothesis, the central device may then provide that intent determination to the input device in operation 480, as well as any commands, queries, or other requests that may be relevant thereto. The input device may then take appropriate action as described above in relation to Figure 3.

[055] According to various aspects of the invention, Figure 6 illustrates a block diagram of an exemplary distributed implementation of the integrated, multi-modal, multi-device natural language voice service environment. As described above, the distributed implementation may also be categorized as a disconnected or peer-to-peer mode that may be employed when a central device in a centralized implementation cannot be reached or otherwise does not meet the needs of the environment. The distributed implementation illustrated in Figure 6 may be generally operate with similar purposes as described above in relation to the centralized implementation (i.e., to ensure that the environment includes a comprehensive model of aggregate knowledge and capabilities of a plurality of devices 610a-n in the environment). Nonetheless, the distributed implementation may operate in a somewhat different manner, in that one or more of the devices 610a-n may be provided with the entire constellation model, or various aspects of the model may be distributed among the plurality of devices 610a-n, or various combinations thereof.

[056] Generally speaking, the plurality of voice-enabled devices 610a- may be coupled to one another by a voice services interface 630, which may include any suitable real or virtual interface (e.g., a common message bus or network interface, a service-oriented abstraction layer, etc.). The various devices 610a-n may therefore operate as cooperative nodes in determining intent for multi-modal natural language utterances received by any one of the devices 610. Furthermore,

Google Exhibit 1019
Page 00749

the devices 610a-n may share knowledge of vocabularies, context, capabilities, and other information, while certain forms of data may be synchronized to ensure consistent processing among the devices 610a-n. For example, because natural language processing components used in the devices 610a-n may vary (e.g., different recognition grammars or speech recognition techniques may exist), vocabulary translation mechanisms, misrecognitions, context variables, criteria values, criteria handlers, and other such information used in the intent determination process should be synchronized to the extent that communication capabilities permit.

[057] By sharing intent determination capabilities, device capabilities, inferencing capabilities, domain knowledge, and other information, decisions as to routing an utterance to a specific one of the devices 610a-n may be made locally (e.g., at an input device), collaboratively (e.g., a device having particular capabilities relevant to the utterance may communicate a request to process the utterance), or various combinations thereof (e.g., the input device may consider routing to secondary devices only when an intent of the utterance cannot be determined). Similarly, partial recognition performed at one or more of the devices 610a-n may be used to determine routing strategies for further intent determination of the utterance. For example, an utterance that contains a plurality of requests relating to a plurality of different domains may be received at an input device that can only determine intent in one of the domains. In this example, the input device may perform partial recognition for the domain associated with the input device, and the partial recognition may also identify the other domains relevant to the utterance for which the input device does not have sufficient recognition information. Thus, the partial recognition performed by the input device may result in identification of other potentially relevant domains and a strategy may be formulated to route the utterance to other devices in the environment that include recognition information for those domains.

[058] As a result, multi-modal natural language inputs, including natural language utterances, may be routed among the various devices 610a-n in order to perform intent determination in a distributed manner. However, as the capabilities and knowledge held by any one of the devices 610a-n may vary, each of the devices 610a-n may be associated with a reliability factor for intent determinations generated by the respective devices 610a-n. As such, to ensure that final intent determinations can be relied upon with a sufficient level of confidence, knowledge may be distributed among the devices 610a-n to ensure that reliability metrics for intent determinations provided by each of the devices 610a-n are commensurable throughout the environment. For

Google Exhibit 1019
Page 00750

example, additional knowledge may be provided to a device having a low intent determination reliability, even when such knowledge results in redundancy in the environment, to ensure commensurate reliability of intent determination environment-wide.

[059] Therefore, in distributed implementations of the integrated voice services environment, utterances may be processed in various ways, which may depend on circumstances at a given time (e.g., system states, system or user preferences, etc.). For example, an utterance may be processed locally at an input device and only routed to secondary devices when an intent determination confidence level falls below a given threshold. In another example, utterances may be routed to a specific device based on the modeling of knowledge and capabilities discussed above. In yet another example, utterances may be flooded among all of the devices in the environment, and arbitration may occur whereby intent determinations may be collated and arbitrated to determine a best guess at intent determination.

[060] Thus, utterances may be processed in various ways, including through local techniques, centralized techniques, distributed techniques, and various combinations thereof. Although many variations will be apparent, Figure 7 illustrates an exemplary method for combined local and distributed processing of multi-modal, natural language inputs in a distributed implementation of the voice service environment, according to various aspects of the invention. In particular, the distributed processing may begin in an operation 710, where a multi-modal natural language input may be received at an input device. The input device may then utilize various natural language processing capabilities associated therewith in an operation 720 to transcribe an utterance contained in the multi-modal input (e.g., using an Automatic Speech Recognizer and associated recognition grammars), and may subsequently determine a preliminary intent of the multi-modal natural language input in an operation 730. It will be apparent that operations 710 through 730 may generally be performed using local intent determination capabilities associated with the input device.

[061] Thereafter, the input device may invoke intent determination capabilities of one or more secondary devices in an operation 740. More particularly, the input device may provide information associated with the multi-modal natural language input to one or more of the secondary devices, which may utilize local intent determination capabilities to attempt to determine intent of the input using techniques as described in relation to Figure 5. It will also be apparent that, in various implementations, the secondary devices invoked in operation 740 may

Google Exhibit 1019
Page 00751

include only devices having intent determination capabilities associated with a specific domain identified as being associated with the input. In any event, the input device may receive intent determinations from the invoked secondary devices in an operation 750, and the input device may then collate the intent determinations received from the secondary devices. The input device may then arbitrate among the various intent determinations, or may combine various ones of the intent determinations (e.g., when distinct secondary devices determine intent in distinct domains), or otherwise arbitrate among the intent determinations to determine a best guess at the intent of the multi-modal natural language input (e.g., based on confidence levels associated with the various intent determinations). Based on the determined intent, the input device may then take appropriate action in an operation 770, such as issuing one or more commands, queries, or other requests to be executed at one or more of the input device or the secondary devices.

[062] Furthermore, in addition to the exemplary implementations described above, various implementations may include a continuous listening mode of operation where a plurality of devices may continuously listen for multi-modal voice-based inputs. In the continuous listening mode, each of the devices in the environment may be triggered to accept a multi-modal input when one or more predetermined events occur. For example, the devices may each be associated with one or more attention words, such as "Phone, <multi-modal request>" for a mobile phone, or "Computer, <multi-modal request>" for a personal computer. When one or more of the devices in the environment recognize the associated attention word, keyword activation may result, where the associated devices trigger to accept the subsequent multi-modal request. Further, where a plurality of devices in a constellation may be listening, the constellation may use all available inputs to increase recognition rates.

[063] Moreover, it will be apparent that the continuous listening mode may be applied in centralized voice service environments, distributed centralized voice service environments, or various combinations thereof. For example, when each device in the constellation has a different attention word, any given device that recognizes an attention word may consult a constellation model to determine a target device or functionality associated with the attention word. In another example, when a plurality of devices in the constellation share one or more attention words, the plurality of devices may coordinate with one another to synchronize information for processing the multi-modal input, such as a start time for an utterance contained therein.

Google Exhibit 1019
Page 00752

[064] Implementations of the invention may be made in hardware, firmware, software, or various combinations thereof. The invention may also be implemented as instructions stored on a machine-readable medium, which may be read and executed by one or more processors. A machine-readable medium may include various mechanisms for storing or transmitting information in a form readable by a machine (e.g., a computing device). For example, a machine-readable storage medium may include read only memory, random access memory, magnetic disk storage media, optical storage media, flash memory devices, and others, and a machine-readable transmission media may include forms of propagated signals, such as carrier waves, infrared signals, digital signals, and others. Further, firmware, software, routines, or instructions may be described in the above disclosure in terms of specific exemplary aspects and implementations of the invention, and performing certain actions. However, it will be apparent that such descriptions are merely for convenience and that such actions in fact result from computing devices, processors, controllers, or other devices executing the firmware, software, routines, or instructions.

[065] Aspects and implementations may be described as including a particular feature, structure, or characteristic, but every aspect or implementation may not necessarily include the particular feature, structure, or characteristic. Further, when a particular feature, structure, or characteristic has been described in connection with an aspect or implementation, it will be understood that such feature, structure, or characteristic may be included in connection with other aspects or implementations, whether or not explicitly described. Thus, various changes and modifications may be made to the preceding description without departing from the scope or spirit of the invention, and the specification and drawings should therefore be regarded as exemplary only, and the scope of the invention determined solely by the appended claims.

Google Exhibit 1019
Page 00753

CLAIMS

What is claimed is:

1. A method for providing an integrated natural language voice services environment, the integrated voice services environment including an input device that receives a multi-modal natural language input, a central device communicatively coupled to the input device, and one or more secondary devices communicatively coupled to the central device, each of the input device, the central device, and the one or more secondary devices having intent determination capabilities for processing multi-modal natural language inputs, the method comprising:

communicating the multi-modal natural language input from the input device to the central device, the central device operable to:

aggregate the intent determination capabilities of the input device and the one or more secondary devices; and

determine an intent of the multi-modal natural language input received at the input device using the aggregated intent determination capabilities;

receiving the determined intent of the multi-modal natural language input at the input device; and

invoking at least one action at one or more of the input device, the central device, or the secondary devices based on the determined intent.

2. The method of claim 1, the intent determination capabilities for a given one of the input device, the central device, or the secondary devices based on at least one of processing power, storage resources, natural language processing capabilities, or local knowledge.

3. The method of claim 1, the central device further operable to:

communicate the multi-modal natural language input to each of the secondary devices, wherein each of the secondary devices determine an intent of the multi-modal natural language input received at the input device using local intent determination capabilities;

receive the intent determined by each of the secondary devices; and

arbitrate among the intent determinations of the secondary devices to determine the intent of the multi-modal natural input.

Google Exhibit 1019
Page 00754

4. The method of claim 3, wherein the central device arbitrates among the intent determinations of the secondary devices based on respective confidence levels associated therewith.

5. The method of claim 4, the central device further operable to weigh the confidence level of the intent determinations of the secondary devices, wherein intent determinations of secondary devices that use multi-pass speech recognition techniques receive a higher weight.

6. The method of claim 1, further comprising sharing knowledge among the input device, the central device, and the secondary devices to update the intent determination capabilities thereof.

7. The method of claim 6, wherein the shared knowledge includes information relating to at least one of vocabularies, translation mechanisms, misrecognitions, context, or states of the devices in the environment.

8. The method of claim 1, the input device operable to:

determine an intent of the multi-modal natural language input received at the input device using local intent determination capabilities; and

invoke the at least one action without communicating the multi-modal natural language input to the central device when the intent determination of the input device meets a predetermined level of confidence.

9. The method of claim 1, wherein the voice services environment operates in a continuous listening mode.

10. A method for providing an integrated natural language voice services environment, the integrated voice services environment including an input device that receives a multi-modal natural language input and one or more secondary devices communicatively coupled to the input device, each of the input device and the one or more secondary devices having intent

Google Exhibit 1019
Page 00755

determination capabilities for processing multi-modal natural language inputs, the method comprising:

determining a preliminary intent of the multi-modal natural language input, wherein the input device determines the preliminary intent using local intent determination capabilities;

communicating the multi-modal natural language input to one or more of the secondary devices, each of which determine an intent of the multi-modal natural language input using local intent determination capabilities;

collating the preliminary intent determination and the intent determinations of the secondary devices;

arbitrating among the collated intent determinations to determine an actionable intent of the multi-modal natural input; and

invoking at least one action at one or more of the input device or the secondary devices based on the determined actionable intent.

11. The method of claim 10, further comprising sharing knowledge among the input device and the secondary devices relating to respective local intent determination capabilities.

12. The method of claim 11, further comprising associating a reliability factor with the local intent determination capabilities of each of the input device and the one or more secondary devices, wherein an aggregation of the reliability factors for the local intent determination capabilities provide commensurable intent determination capabilities throughout the environment.

13. The method of claim 11, wherein the shared knowledge includes information relating to at least one of processing power, storage resources, natural language processing capabilities, or knowledge associated with the input device and the secondary devices.

14. The method of claim 10, further comprising sharing knowledge among the input device and the secondary devices to update the intent determination capabilities thereof.

Google Exhibit 1019
Page 00756

15. The method of claim 14, wherein sharing the knowledge includes transmitting knowledge established on the input device to the one or more secondary devices.

16. The method of claim 14, wherein sharing the knowledge includes transmitting knowledge established on one of the secondary devices to the input device and to the secondary devices other than the device where the knowledge was established.

17. The method of claim 14, wherein the shared knowledge includes information relating to at least one of vocabularies, translation mechanisms, misrecognitions, context, or states of the devices in the environment.

18. The method of claim 10, wherein the voice services environment operates in a continuous listening mode.

19. A system for providing an integrated natural language voice services environment, the system comprising:

an input device that receives a multi-modal natural language input;

one or more secondary devices communicatively coupled to the input device, each of the input device and the one or more secondary devices having intent determination capabilities for processing multi-modal natural language inputs; and

a constellation model accessible to each of the input device and the one or more secondary devices, the constellation model describing the intent determination capabilities of the input device and the one or more secondary devices, the multi-modal natural language input routed for processing at one or more of the input device or the secondary devices to determine an intent thereof based on the intent determination capabilities described in the constellation model.

20. The system of claim 19, wherein the constellation model describes, for each of the input device and the one or more secondary devices, one or more of processing power, storage resources, natural language processing capabilities, or knowledge associated therewith.

Google Exhibit 1019
Page 00757

21. The system of claim 19, wherein the constellation model dynamically selecting one of a centralized model or a distributed model for operating the environment.

22. The system of claim 21, the centralized model designating one of the secondary devices as a central device, the central device operable to:

aggregate the intent determination capabilities of the input device and the one or more secondary devices; and

determine the intent of the multi-modal natural language input using the aggregated intent determination capabilities.

23. The system of claim 22, wherein the constellation model selects the distributed model when the multi-modal natural language cannot be communicated to the central device.

24. The system of claim 23, wherein selecting the distributed model causes the input device and the one or more secondary devices to:

share knowledge relating to respective local intent determination capabilities; and

operate as cooperative nodes to determine the intent of the multi-modal natural language input using the shared knowledge relating to local intent determination capabilities.

25. The system of claim 24, wherein operating as cooperative nodes includes determining whether the intent of the multi-modal natural language input should be determined by the input device, by one or more of the secondary devices, or by each of the input device and the one or more secondary devices.

26. The system of claim 19, further comprising at least one domain agent associated with the input device, wherein the at least one domain agent bootstraps information relating to a context associated with the domain agent from one or more of the secondary devices.

27. The system of claim 19, wherein the voice services environment operates in a continuous listening mode.

Google Exhibit 1019
Page 00758

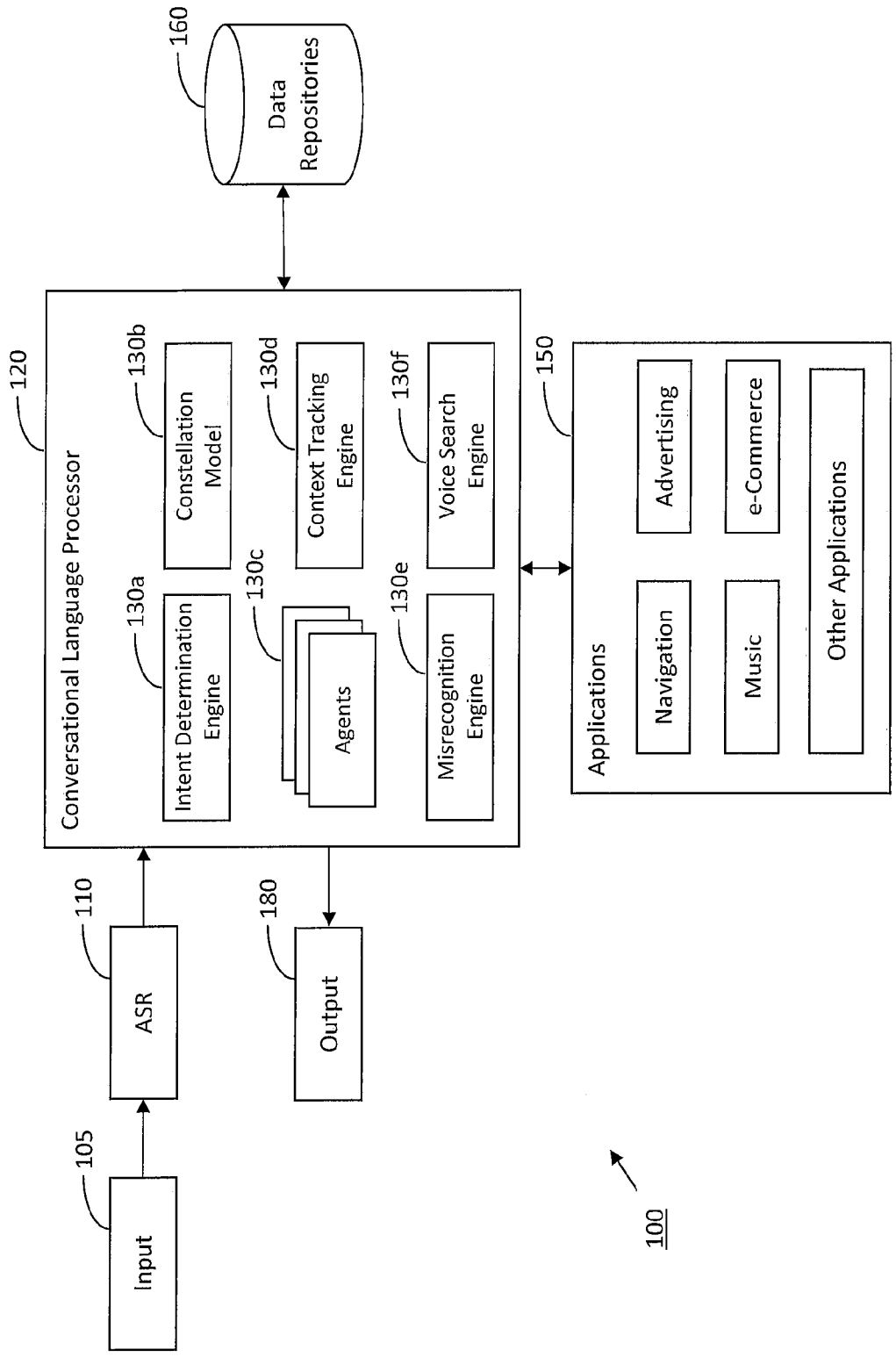

160
Data Repositories
120
Conversational Language Processor
130b
Constellation Model
130d
Context Tracking Engine
130f
Voice Search Engine
150
130a
Intent Determination Engine
130c
Agents
130e
Misrecognition Engine
Applications
Advertising
Navigation
e-Commerce
Music
Other Applications
110
ASR
180
Output
105
Input
100

Figure 1

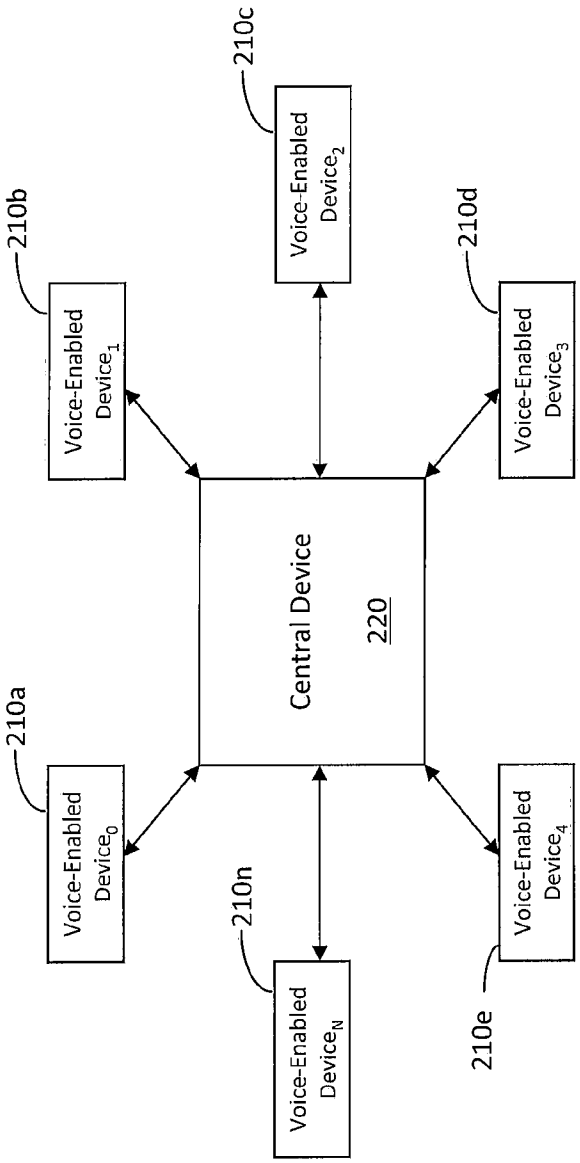

210c
Voice-Enabled Device$_2$
210b
Voice-Enabled Device$_1$
210d
Voice-Enabled Device$_3$
Central Device
220
210a
Voice-Enabled Device$_0$
210n
Voice-Enabled Device$_N$
Voice-Enabled Device$_4$
210e

Figure 2

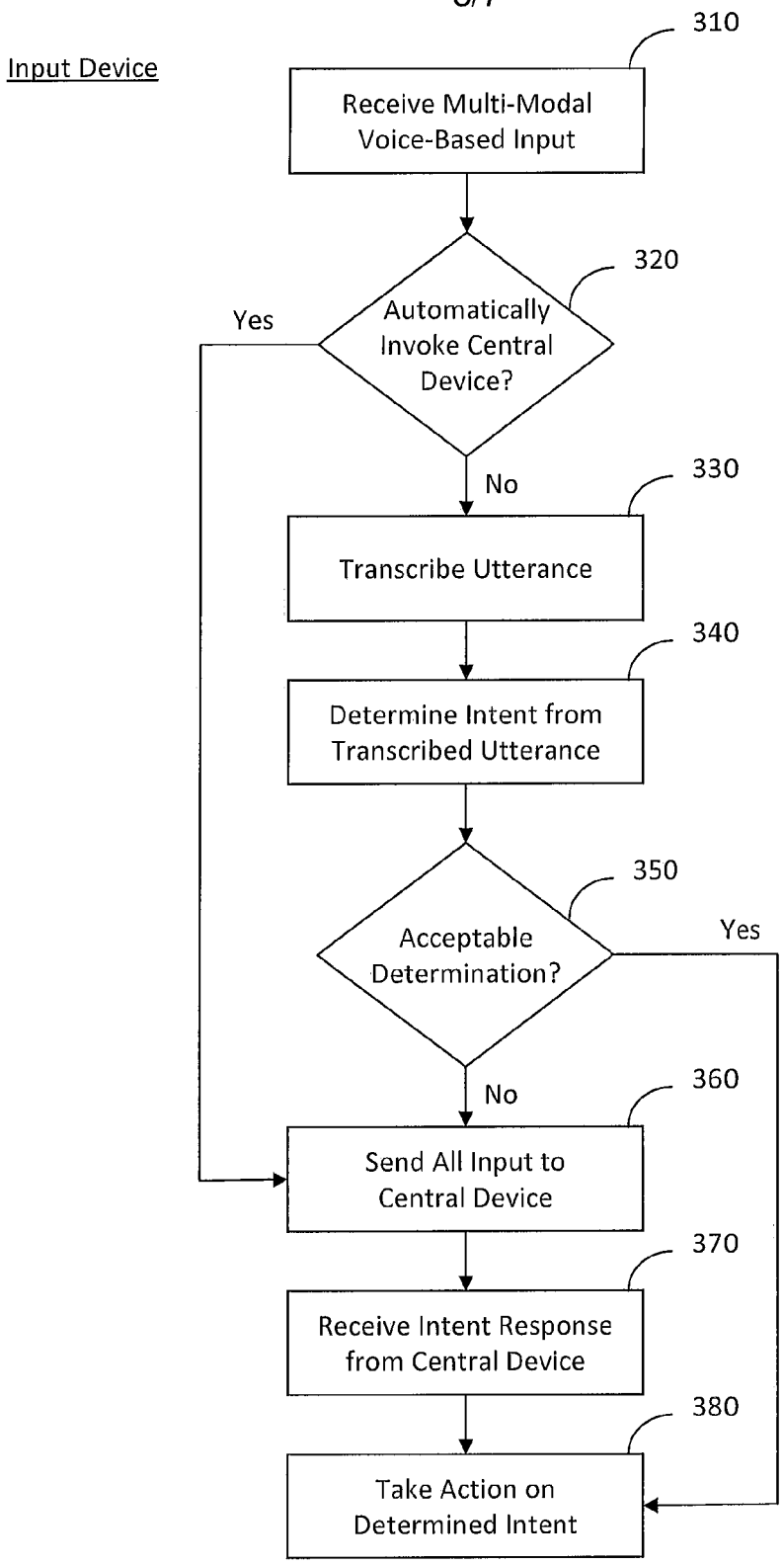

Input Device
310
Receive Multi-Modal Voice-Based Input
320
Automatically Invoke Central Device?
Yes
No
330
Transcribe Utterance
340
Determine Intent from Transcribed Utterance
350
Acceptable Determination?
Yes
No
360
Send All Input to Central Device
370
Receive Intent Response from Central Device
380
Take Action on Determined Intent

Figure 3

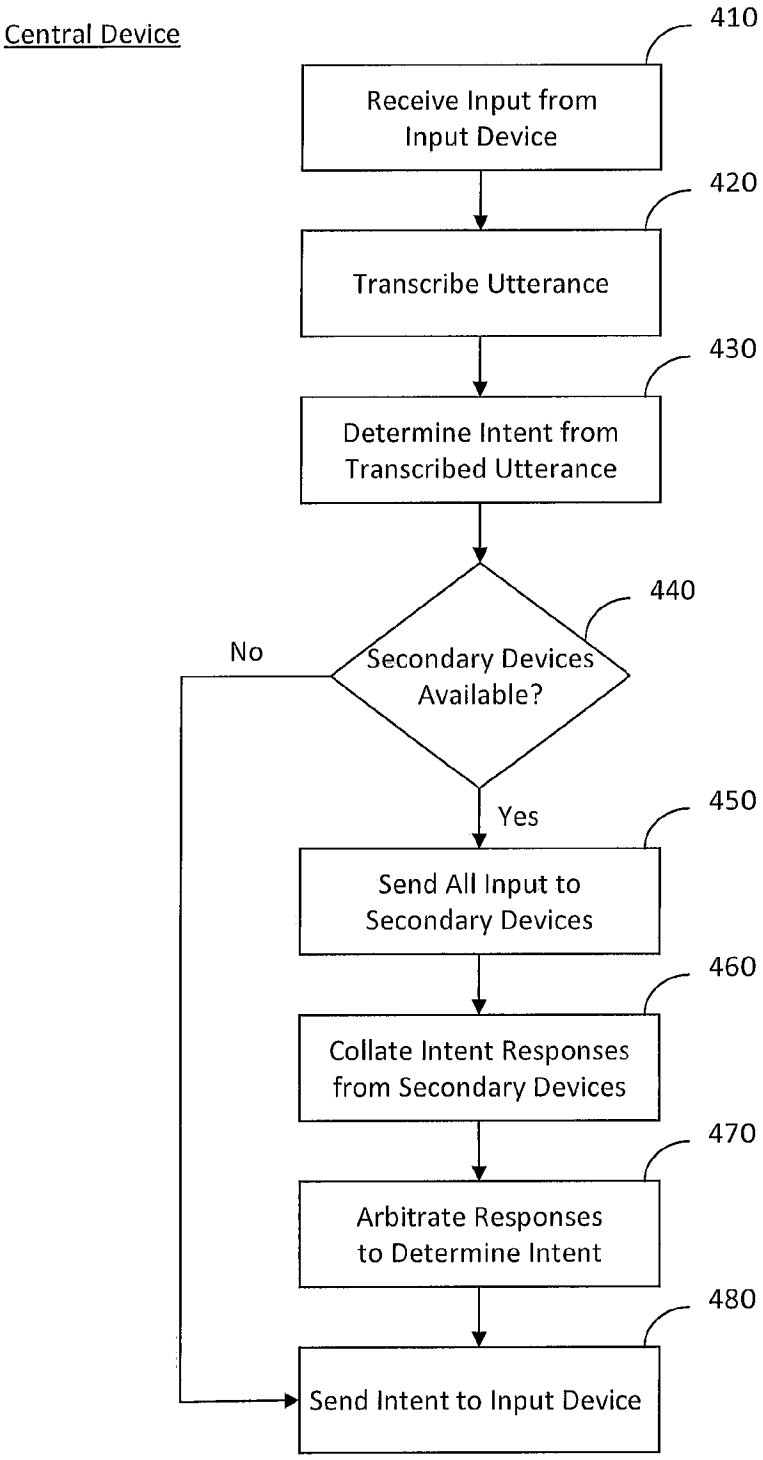

Central Device
410
Receive Input from Input Device
420
Transcribe Utterance
430
Determine Intent from Transcribed Utterance
440
Secondary Devices Available?
No
Yes
450
Send All Input to Secondary Devices
460
Collate Intent Responses from Secondary Devices
470
Arbitrate Responses to Determine Intent
480
Send Intent to Input Device

Figure 4

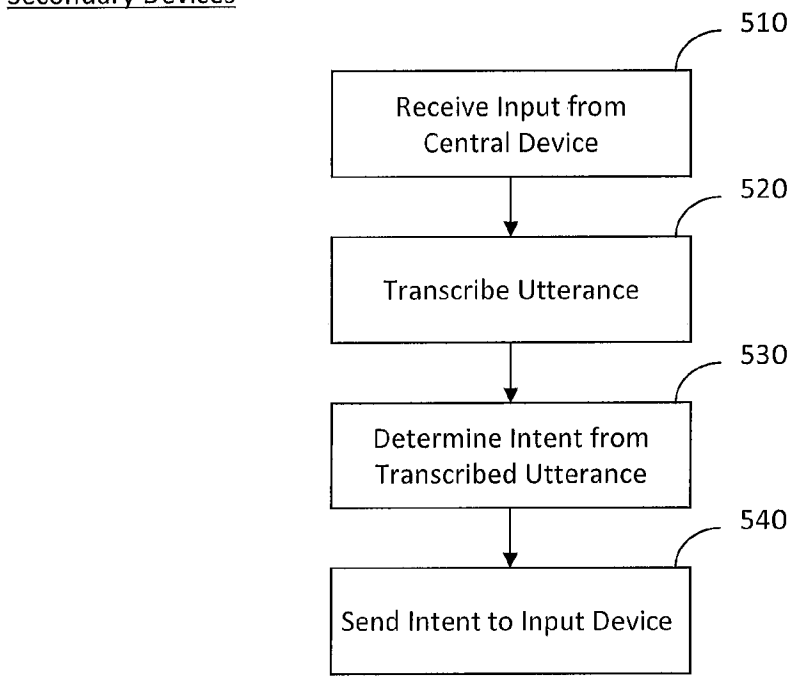

Secondary Devices
510
Receive Input from Central Device
520
Transcribe Utterance
530
Determine Intent from Transcribed Utterance
540
Send Intent to Input Device

Figure 5

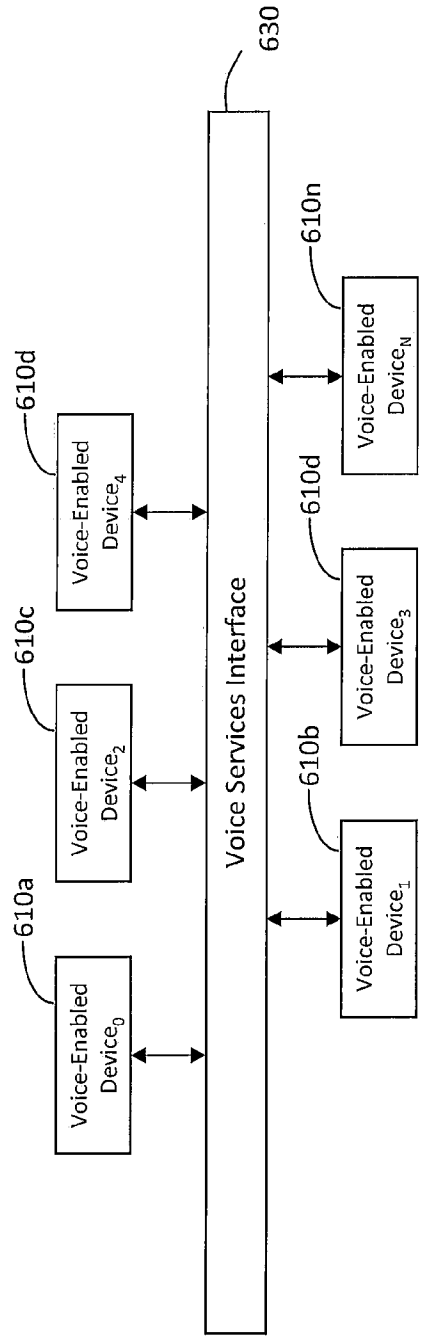

630
Voice Services Interface
610a
Voice-Enabled Device$_0$
610b
Voice-Enabled Device$_1$
610c
Voice-Enabled Device$_2$
610d
Voice-Enabled Device$_3$
610d
Voice-Enabled Device$_4$
610n
Voice-Enabled Device$_N$

Figure 6

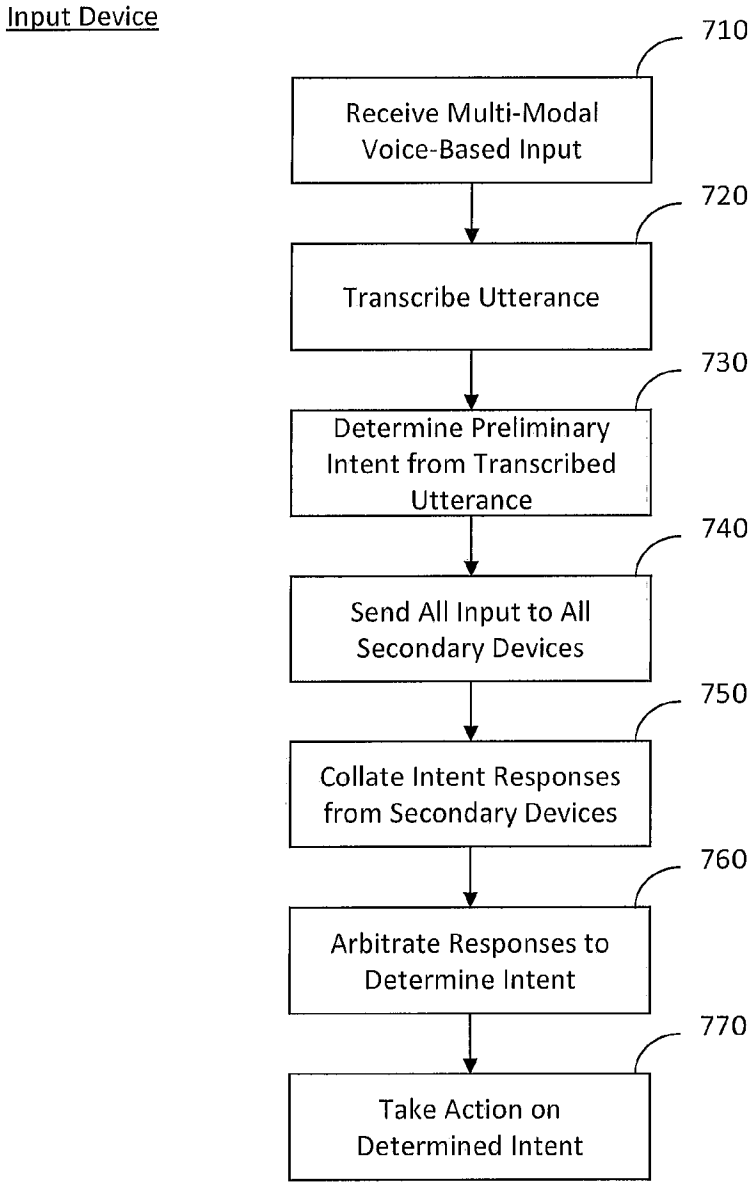

Input Device
710
Receive Multi-Modal Voice-Based Input
720
Transcribe Utterance
730
Determine Preliminary Intent from Transcribed Utterance
740
Send All Input to All Secondary Devices
750
Collate Intent Responses from Secondary Devices
760
Arbitrate Responses to Determine Intent
770
Take Action on Determined Intent

Figure 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 08/69524

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 13/12 (2008.04)
USPC - 710/65
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 710/65

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 710/11, 14, 37, 50, 73; 370/353, 458, 448; 379/907-908; 704/200, 246, 251-252 (see text terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (PGPB, USPT, USOC, EPAB, JPAB); DialogPRO; Google (incl. Patents, Scholar) - natural language, multi-modal, central, input, device, client, server, process, distributed, local, distributed system, local, context, speech

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 7,376,645 B2 (BERNARD) 20 May 2008 (20.05.2008), Abstract, FIGS.1-3; and Col 1, ln 19-30; Col 1, ln 35 to Col 2, ln 2; Col 2, ln 21 to Col 4, ln 16; Col 4, ln 21-56, 60; Col 5, ln 10; Col 6, ln 1-5, 7, 22-23; Col 6, ln 37 to Col 7, ln 4; Col 7, ln 13-26; Col 7, ln 52 to Col 10, ln 37	1-27
A	US 2004/0025115 A1 (SIENEL et al.) 5 February 2004 (05.02.2004)	1-27
A	US 2007/0050191 A1 (WEIDER et al.) 1 March 2007 (01.03.2007)	1-27
A	US 2004/0166832 A1 (PORTMAN et al.) 26 August 2004 (26.08.2004)	1-27
A	US 6,615,172 B1 (BENNETT et al.) 2 September 2003 (02.09.2003)	1-27

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 October 2008 (04.10.2008)	09 OCT 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 August 2010 (26.08.2010)

PCT

(10) International Publication Number
WO 2010/096752 A1

(51) International Patent Classification:
G10L 15/00 (2006.01)

(21) International Application Number:
PCT/US2010/024895

(22) International Filing Date:
22 February 2010 (22.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/389,678 20 February 2009 (20.02.2009) US

(71) Applicant *(for all designated States except US)*: **VOICEBOX TECHNOLOGIES, INC.** [US/US]; 11980 NE 24th Street, Suite 100, Bellevue, WA 98005 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BALDWIN, Larry** [US/US]; 23125 SE 243rd Place, Maple Valley, WA 98038 (US). **WEIDER, Chris** [US/US]; 10757 Greenwood Avenue North #220, Seattle, WA 98133 (US).

(74) Agents: **ALI, Syed, Jafar** et al.; Pillsbury Winthrop Shaw Pittman LLP, P.O. Box 10500, McLean, VA 22102 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: SYSTEM AND METHOD FOR PROCESSING MULTI-MODAL DEVICE INTERACTIONS IN A NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT

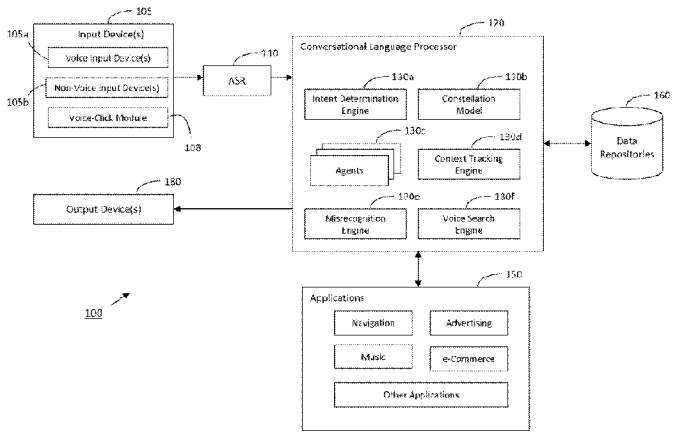


Figure 1

(57) **Abstract**: A system and method for processing multi-modal device interactions in a natural language voice services environment may be provided. In particular, one or more multi-modal device interactions may be received in a natural language voice services environment that includes one or more electronic devices. The multi-modal device interactions may include a non-voice interaction with at least one of the electronic devices or an application associated therewith, and may further include a natural language utterance relating to the non-voice interaction. Context relating to the non-voice interaction and the natural language utterance may be extracted and combined to determine an intent of the multi-modal device interaction, and a request may then be routed to one or more of the electronic devices based on the determined intent of the multi-modal device interaction.

WO 2010/096752 A1

SYSTEM AND METHOD FOR PROCESSING MULTI-MODAL DEVICE INTERACTIONS IN A NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT

CROSS-REFERENCE TO RELATED APPLICATIONS

[001] This application claims the benefit of U.S. Patent Application Serial No. 12/389,678, entitled "SYSTEM AND METHOD FOR PROCESSING MULTI-MODAL DEVICE INTERACTIONS IN A NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT", filed February 20, 2009, which is hereby incorporated by reference in its entirety.

FIELD OF THE INVENTION

[002] The invention relates to an integrated natural language voice services environment for processing multi-modal interactions with or one or more devices and/or applications, wherein the multi-modal interactions may provide additional context for cooperatively interpreting and otherwise processing a natural language utterance accompanying the multi-modal interactions.

BACKGROUND OF THE INVENTION

[003] As technology has progressed in recent years, consumer electronic devices have emerged to become nearly ubiquitous in the everyday lives of many people. To meet the increasing demand that has resulted from growth in the functionality and mobility of mobile phones, navigation devices, embedded devices, and other such devices, many devices offer a wealth of features and functions in addition to core applications. Greater functionality also introduces trade-offs, however, including learning curves that often inhibit users from fully exploiting all of the capabilities of their electronic devices. For example, many existing electronic devices include complex human to machine interfaces that may not be particularly user-friendly, which can inhibit mass-market adoption for many technologies. Moreover, cumbersome interfaces often result in otherwise desirable features being difficult to find or use (e.g., because of menus that are complex or otherwise tedious to navigate). As such, many users tend not to use, or even know about, many of the potential capabilities of their devices.

[004] As such, the increased functionality of electronic devices often tends to be wasted, as market research suggests that many users only use only a fraction of the features or applications available on a given device. Moreover, in a society where wireless networking and broadband access are increasingly prevalent, consumers tend to naturally desire seamless

Google Exhibit 1019
Page 00768

mobile capabilities from their electronic devices. Thus, as consumer demand intensifies for simpler mechanisms to interact with electronic devices, cumbersome interfaces that prevent quick and focused interaction become an important concern. Nevertheless, the ever-growing demand for mechanisms to use technology in intuitive ways remains largely unfulfilled.

[005] One approach towards simplifying human to machine interactions in electronic devices has included the use of voice recognition software, which has the potential to enable users to exploit features that would otherwise be unfamiliar, unknown, or difficult to use. For example, a recent survey conducted by the Navteq Corporation, which provides data used in a variety of applications such as automotive navigation and web-based applications, demonstrates that voice recognition often ranks among the features most desired by consumers of electronic devices. Even so, existing voice user interfaces, when they actually work, still require significant learning on the part of the user.

[006] For example, many existing voice user interface only support requests formulated according to specific command-and-control sequences or syntaxes. Furthermore, many existing voice user interfaces cause user frustration or dissatisfaction because of inaccurate speech recognition. Similarly, by forcing a user to provide pre-established commands or keywords to communicate requests in ways that a system can understand, existing voice user interfaces do not effectively engage the user in a productive, cooperative dialogue to resolve requests and advance a conversation towards a satisfactory goal (e.g., when users may be uncertain of particular needs, available information, device capabilities, etc.). As such, existing voice user interfaces tend to suffer from various drawbacks, including significant limitations on engaging users in a dialogue in a cooperative and conversational manner.

[007] Additionally, many existing voice user interfaces fall short in utilizing information distributed across different domains, devices, and applications in order to resolve natural language voice-based inputs. Thus, existing voice user interfaces suffer from being constrained to a finite set of applications for which they have been designed, or to devices on which they reside. Although technological advancement has resulted in users often having several devices to suit their various needs, existing voice user interfaces do not adequately free users from device constraints. For example, users may be interested in services associated with different applications and devices, but existing voice user interfaces tend to restrict users from accessing the applications and devices as they see fit. Moreover, users typically can only practicably carry

Google Exhibit 1019
Page 00769

a finite number of devices at any given time, yet content or services associated with users' other devices currently being used may be desired in various circumstances.

[008] Accordingly, although users tend to have varying needs, where content or services associated with different devices may be desired in various contexts or environments, existing voice technologies tend to fall short in providing an integrated environment in which users can request content or services associated with virtually any device or network. As such, constraints on information availability and device interaction mechanisms in existing voice services environments tend to prevent users from experiencing technology in an intuitive, natural, and efficient way. For instance, when a user wishes to perform a given function using a given electronic device but does not necessarily know how to go about performing the function, the user typically cannot engage in a multi-modal interaction with the device to simply utter words in natural language to request the function.

[009] Furthermore, relatively simple functions can often be tedious to perform using electronic devices that do not have voice recognition capabilities. For example, purchasing new ring-tones for a mobile phone tends to be a relatively straightforward process, but users must typically navigate several menus and press many different buttons in order to complete the process. As such, it becomes apparent that interaction with electronic devices could be far more efficient if users were able to use natural language to exploit buried or otherwise difficult to use functionality. Existing systems suffer from these and other problems.

SUMMARY OF THE INVENTION

[010] According to one aspect of the invention, a system and method for processing multi-modal device interactions in a natural language voice services environment may be provided. In particular, one or more multi-modal interactions may be received in a natural language voice services environment that includes one or more electronic devices. The multi-modal device interactions may include a user engaging in a non-voice interaction with one or more of the electronic devices or applications associated with the devices, while also providing a natural language utterance in relation to the non-voice interaction. For example, the non-voice device interaction may comprise the user selecting a particular segment, item, data, point of focus, or attention focus, or otherwise engaging in one or more unique and distinguishable interactions with the electronic devices or applications associated therewith. As such, context may be

Google Exhibit 1019
Page 00770

extracted from the natural language utterance, and the non-voice device interaction may provide additional context for the natural language utterance. The context of the utterance and the non-voice device interaction may then be combined to determine an intent of the multi-modal device interaction, wherein one or more of the electronic devices may process a request based on the intent of the multi-modal device interaction.

[011] According to one aspect of the invention, at least one of the electronic devices may include an input device configured to receive voice-based inputs. In one implementation, the voice-based input device may be signaled to capture the natural language utterance in response to detecting the non-voice interaction with the one or more electronic devices or applications. Furthermore, the natural language voice services environment may include one or more listeners established for the electronic devices and the associated applications, wherein the listeners may be configured to detect the non-voice interaction with the electronic devices or applications. As such, information relating to the non-voice interaction and the accompanying natural language utterance may be aligned to enable cooperative processing of the utterance and the non-voice device interaction.

[012] According to one aspect of the invention, at least one transaction lead may be generated based on the intent of the multi-modal device interaction. For example, a further multi-modal device interaction may be received, wherein the further multi-modal device interaction may relate to the transaction lead generated for the first multi-modal device interaction. At least one request may then be routed to one or more of the electronic devices based on an intent determined for the further multi-modal device interaction, whereby a transaction click-through may be processed in response to receiving the device interaction relating to the generated transaction lead. For example, the transaction lead may include an advertisement or a recommendation that is selected based on the intent of the original multi-modal device interaction, while the further multi-modal device interaction may include a user selecting the advertisement or recommendation. Thus, the selection of the advertisement or recommendation may be considered a transaction click-through, which may generate revenue for a particular entity (e.g., a provider of the natural language voice services environment).

[013] Other objects and advantages of the invention will be apparent based on the following drawings and detailed description.

Google Exhibit 1019
Page 00771

BRIEF DESCRIPTION OF THE DRAWINGS

[014] Figure 1 illustrates a block diagram of an exemplary system for processing multi-modal device interactions in a natural language voice services environment, according to various aspects of the invention.

[015] Figure 2 illustrates a block diagram of an exemplary method for synchronizing multi-modal devices in a natural language voice service environment, according to various aspects of the invention.

[016] Figure 3 illustrates a flow diagram of an exemplary method for processing multi-modal device interactions in a natural language voice service environment, according to various aspects of the invention.

[017] Figure 4 illustrates a flow diagram of an exemplary method for processing multi-modal device interactions to generate one or more transaction leads in a natural language voice service environment, according to various aspects of the invention.

DETAILED DESCRIPTION

[018] According to various aspects of the invention, Figure 1 illustrates a block diagram of an exemplary system 100 for processing multi-modal device interactions in a natural language voice services environment. As will be apparent from the further description to be provided herein, the system 100 illustrated in Figure 1 may include an input device 105, or a combination of input devices 105, which enable users to interact with the system 100 in a multi-modal manner. In particular, the system 100 may include various natural language processing components, including at least a voice-click module 108, which can collectively process a user's multi-modal interactions with one or more input devices 105. For example, in one implementation, the input devices 105 may include any suitable combination of at least one voice input device 105a (e.g., a microphone) and at least one non-voice input device 105b (e.g., a mouse, touch-screen display, wheel selector, etc.). As such, the input devices 105 may include any suitable combination of electronic devices having mechanisms for receiving both voice-based and non-voice-based inputs (e.g., a microphone coupled to one or more of a telematics device, personal navigation device, mobile phone, VoIP node, personal computer, media device, embedded device, server, or other electronic device). As such, the system 100 may enable users to engage in multi-modal conversational interactions with the one or more

Google Exhibit 1019
Page 00772

electronic input devices 105 or applications associated with the electronic devices 105, wherein the system 100 may process the device interactions in a free-form and cooperative manner suitable for routing tasks or otherwise resolving requests.

[019] As indicated above, in one implementation, the system may include various natural language processing components that can support free-form utterances and/or other forms of device interactions, which may liberate users from restrictions relating to the manner of formulating commands, queries, or other requests. As such, users may interact with the input devices 105 using any manner of speaking into the voice input device 105a or otherwise interacting with the non-voice input devices 105b in order to request content or services available in the system 100. For instance, a user may request any content or services that may be available in the system 100 by providing a natural language utterance to the voice input device 105a. In one implementation, the utterance may then be processed using techniques described in U.S. Patent No. 7,398,209, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," issued July 8, 2008, and U.S. Patent Application 10/618,633, entitled "Mobile Systems and Methods for Responding to Natural Language Speech Utterance," filed June 15, 2003, the disclosures of which are hereby incorporated by reference in their entirety. In addition, the user may interact with one or more of the non-voice input devices 105b to provide further context or other information relating to the utterance and/or the requested content or services.

[020] In one implementation, the system 100 may be coupled to various other systems that include additional multi-modal devices, with the other systems having natural language processing capabilities similar to those shown in Figure 1. The system 100 may therefore provide an interface to a multi-device environment in which users may request content or services available through the various additional devices in the environment. For example, in one implementation, the system 100 may include a constellation model 130b that provides knowledge relating to content, services, applications, intent determination capabilities, and other features available through the other systems and devices in the environment. For example, in one implementation, the system 100 may interact with devices, applications, or other systems in the environment to cooperatively resolve requests, as described in co-pending U.S. Patent Application 12/127,343, entitled "System and Method for an Integrated, Multi-Modal, Multi-Device Natural Language Voice Services Environment," filed May 27, 2008, the

Google Exhibit 1019
Page 00773

disclosure of which is hereby incorporated by reference in its entirety. For example, the multi-device environment may share information among various systems and devices to provide a cooperative environment for resolving requests, wherein the shared information may relate to aspects such as device capabilities, context, prior interactions, domain knowledge, short-term knowledge, long-term knowledge, and cognitive models, among other things.

[021] As indicated above, the system 100 illustrated in Figure 1 may include, among other things, one or more electronic input devices 105 that collectively provide an interface (or combination of interfaces) for receiving one or more multi-modal device interactions from a user, wherein the device interactions include at least a user-spoken utterance. Although the implementation illustrated in Figure 1 includes a distinct voice input device 105a and non-voice input device 105b, it will be apparent that in one or more implementations the voice input device 105a and the non-voice input device 105b may be components of the same or separate devices. For example, the input devices 105 may include a microphone coupled to a mobile phone (i.e., the voice input device 105a), and may further include one or more buttons, selectable displays, wheel selectors, or other components coupled to the mobile phone (i.e., the non-voice input devices 105b). In another example, the input devices 105 may include a combination of a microphone coupled to a telematics device (i.e., the voice input device 105a), and may further include buttons, a touch-screen display, a track wheel, or other non-voice input devices 105b coupled to a media player that is communicatively coupled to, yet distinct from, the telematics device. Thus, the input devices 105 may include any suitable combination of communicatively coupled electronic devices that includes at least one input device for receiving natural language utterance inputs and at least one input device for receiving multi-modal non-voice inputs.

[022] In one implementation, a voice-click module 108 communicatively coupled to the one or more input devices 105 may enable cooperative processing of multi-modal device interactions received at the voice input device 105a and one or more of the non-voice input devices 105b. For example, the voice-click module 108 may provide the system 100 with information that can be used to process a natural language utterance received via the voice input device 105a in view of one or more non-voice device interactions received via the non-voice input devices 105b. The voice-click module 108 therefore enables the user to interact with the various input devices 105 in an intuitive and free-form manner, whereby the user may provide various types

Google Exhibit 1019
Page 00774

of information to the system 100 when seeking to initiate action, retrieve information, or otherwise request content or services available in the system 100.

[023] The voice input device 105a may comprise any appropriate device, or combination of devices, which have capabilities for receiving natural language utterances or other forms of spoken input. For example, in one implementation, the voice input device 105a may include a directional microphone, an array of microphones, or other devices capable of creating encoded speech. In one implementation, the voice input device 105a may be configured to maximize fidelity of the encoded speech. For example, the voice input device 105a may be configured to maximize gain in a direction of the user, cancel echoes, null point noise sources, perform variable rate sampling, filter environmental noise or background conversations, or use other techniques to maximize the fidelity of the encoded speech. As such, the voice input device 105a may create the encoded speech in a manner tolerant of noise or other factors that could otherwise interfere with the system 100 accurately interpreting natural language utterances.

[024] The non-voice input device 105b may comprise any appropriate device, or combination of devices, which have capabilities for supporting non-voice device interactions. For example, in one implementation, the non-voice input device 105b may include a combination stylus and touch-screen or tablet interface, a BlackBerry® wheel selector, an iPod® click wheel, a mouse, a keypad, buttons, or any other devices that supports distinguishable non-voice device interactions. The user may therefore use the non-voice input devices 105b to make data selections or identify a point of focus (or attention focus) to be processed in connection with a related natural language utterance provided via the voice input device 105a. For example, the user may point a stylus at a specific segment of a touch-screen display, highlight text using a mouse, click a button, interact with an application, or otherwise engage in any suitable device interaction for selecting data or otherwise identifying a point of focus (i.e., voice-activating or "voice-clicking" the selected data and/or the identified point of focus).

[025] Furthermore, in addition to being usable for making data selections, identifying points of focus, or otherwise activating data to be interpreted in association with one or more utterances, the user may further use the non-voice input devices 105b to engage in specialized device interactions that have meaning in the system 100. For example, the specialized device interactions (which may be referred to as "clicks" or "voice-clicks") may include clicks lasting a given duration, clicks continuously held for a given duration, clicks made in a predetermined

Google Exhibit 1019
Page 00775

sequence, or any other interaction or sequence of interactions that the input devices 105 and/or the voice-click module 108 can identify, detect, or otherwise distinguish.

[026] In one implementation, the specialized device interactions may be associated with one or more actions, queries, commands, tasks, or other requests associated with applications or services available in the system 100. In one implementation, the specialized device interactions may further include one or more actions, queries, commands, tasks, or other requests associated with any of various devices deployed in the multi-device environment, as described in the above-referenced co-pending U.S. Patent Application 12/127,343, entitled "System and Method for an Integrated, Multi-Modal, Multi-Device Natural Language Voice Services Environment," filed May 27, 2008. For example, a distinct sequence of clicking a stylus on a particular segment or item displayed on a touch-screen display may be defined as a specialized device interaction or voice-click for initiating a telephone call on a mobile phone, calculating a route on a navigation device, purchasing a song to a media player, or another type of request.

[027] Thus, the voice-click module 108 coupled to the input device 105 may continually monitor the user's interactions with the non-voice input device 105b to detect occurrences of at least one non-voice device interaction, which may be referred to herein as a "voice-click." The detected voice-click may therefore provide further context for processing a multi-modal device interaction, which may include the at least one voice-click and one or more natural language utterances, each of which may provide context for task specification. Thus, the voice-click may generally signal the system 100 that a current utterance or other voice-based input is to be processed together with a current interaction with one or more of the devices 105. For example, in one implementation, the current device interaction may include the user selecting, highlighting, or otherwise identifying a particular point of attention, object, or other item associated with one or more of the devices 105. As such, the current device interaction may provide context for sharpening recognition, interpretation, and understanding of the accompanying utterance, and moreover, the current utterance may provide information to enhance the context provided by the accompanying device interaction.

[028] In one implementation, the voice-click module 108 may determine the various voice-click interactions to be detected based on particular characteristics of the non-voice input devices 105b (e.g., the voice-click interactions may include distinguishable interactions that the non-voice input devices 105b support). For example, a multi-touch display typically includes a

Google Exhibit 1019
Page 00776

touch-screen display device configured to support various distinguishable gestures for interacting with information displayed therein (e.g., a user may zoom in, zoom out, rotate, or otherwise control graphical information displayed on a multi-touch screen using specific gestures or other interaction techniques). Thus, in one example, the non-voice input devices 105b may include a multi-touch display, in which case the voice-click module 108 may be configured to detect an occurrence of a voice-click when a user engages in one or more of the distinguishable gestures supported by the non-voice multi-touch display 105b.

[029] In one implementation, a user may customize or otherwise modify the voice-click interactions to be detected by the voice-click module 108. In particular, the specific device interactions detected by the voice-click module 108 may be removed or modified, or new device interactions may be added. As such, the voice-click device interactions detected by the voice-click module 108 may include any appropriate interaction or combination of interactions that the non-voice input devices 105b and/or the voice-click module 108 can distinguish.

[030] When the voice-click module 108 detects a user's engaging in a voice-click device interaction, the voice-click module 108 may extract context information associated with the voice-click device interaction for voice-activation. In particular, the voice-click module 108 may identify information relating to a segment, item, point of focus, attention focus, or other data selected by the user, or otherwise identify information relating to a particular device interaction or sequence of device interactions engaged in by the user. The voice-click module 108 thus extracts the identified information relating to the detected voice-click, which may be used as context information to be associated with one or more prior, contemporaneous, or subsequent natural language utterances.

[031] Thus, in response to the voice-click module 108 detecting a voice-click (e.g., selection of an icon, a section of text, specific coordinates on a map display, or other information), the voice-click module 108 may signal the system 100 to use a natural language utterance voice input (which may be received via the voice input device 105a) as further context for determining an action, query, command, task, or other request to perform in order to service the detected voice-click. As such, the various natural language processing components in the system 100 may use the combined context of the voice-click and the accompanying natural language utterance to determine the intent of the voice-click device interaction, and to

Google Exhibit 1019
Page 00777

appropriately route one or more actions, queries, commands, tasks, or other requests to any of the various devices deployed in the multi-device environment.

[032] For instance, in one implementation, the multi-device environment may include a voice-enabled navigation device. Thus, an exemplary voice-click device interaction may include a user touching a stylus to a specific intersection displayed on a touch-screen display 105b associated with the voice-enabled navigation device, while also providing an utterance such as "What restaurants are around here?" into a microphone 105a. In this example, the voice-click module 108 may extract information relating to the voice-clicked intersection, which may be used as context for processing the accompanying utterance (i.e., the selected intersection may provide context for interpreting "around here," as opposed to the user's current location or some other meaning). Moreover, as indicated above, the voice input may be used as additional context for determining task specification. Thus, the utterance may be further processed for recognition and conversational interpretation using the various natural language processing components of the system 100, as described in greater detail below.

[033] In one implementation, an Automatic Speech Recognizer (ASR) 110 may generate one or more preliminary interpretations of the utterance received via the voice input device 105a. For example, the ASR 110 may recognize syllables, words, phrases, or other acoustic characteristics of the utterance using one or more dynamically adaptable recognition grammars. In one implementation, the dynamic recognition grammars may be used to recognize a stream of phonemes using phonetic dictation based on one or more acoustic models (e.g., as described in co-pending U.S. Patent Application 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed August 5, 2005, the disclosure of which is hereby incorporated by reference in its entirety).

[034] In one implementation, the ASR 110 may be configured to perform multi-pass speech recognition, where a primary speech recognition engine may generate a primary transcription of the utterance (e.g., using a large list dictation grammar), and may subsequently request one or more secondary transcriptions from one or more secondary speech recognition engines (e.g., using a virtual dictation grammar having decoy words for out-of-vocabulary words). In one implementation, the primary speech recognition engine may request the secondary transcriptions based on a confidence level for the primary transcription.

Google Exhibit 1019
Page 00778

[035] The recognition grammars employed in the ASR 110 may include various vocabularies, dictionaries, syllables, words, phrases, or other information for recognizing utterances. In one implementation, information contained in the recognition grammars may be dynamically optimized to improve a likelihood of accurate recognition for a given utterance (e.g., following an incorrect interpretation of a word or phrase, the incorrect interpretation may be removed from the grammar to reduce a likelihood of repeating the incorrect interpretation). Additionally, various forms of knowledge can be used to continually optimize the information contained in the recognition grammars on a dynamic basis. For example, the system 100 may have knowledge that includes environmental knowledge (e.g., peer-to-peer affinities, capabilities of the various devices in the environment, etc.), historical knowledge (e.g., frequent requests, prior context, etc.), or short term shared knowledge relating to a current conversational dialogue or interaction, among other types of knowledge.

[036] In one implementation, the information in the recognition grammars may be further optimized according to context or application-specific domains. In particular, similar utterances may be interpreted differently depending on a context to which the utterance relates, including navigation, music, movies, weather, shopping, news, languages, temporal or geographic proximities, or other contexts or domains. For example, an utterance containing the word "traffic" may be subject to different interpretations depending on whether the context relates to navigation (i.e., road conditions), music (i.e., the 1960's rock band), or movies (i.e., the film directed by Steven Soderbergh). Accordingly, the ASR 110 may use various techniques to generate a preliminary interpretation of the natural language utterance, such as those described in the above-referenced co-pending U.S. Patent Applications and/or co-pending U.S. Patent Application 11/513,269, entitled "Dynamic Speech Sharpening," filed August 31, 2006, the disclosure of which is hereby incorporated by reference in its entirety.

[037] As such, the ASR 110 may provide one or more preliminary interpretations of the natural language utterance included in the voice-click to a conversational language processor 120. The conversational language processor 120 may include various natural language processing components collectively configured to model human-to-human conversations or interactions. For example, the conversational language processor 120 may include, among other things, an intent determination engine 130a, a constellation model 130b, one or more domain agents 130c, a context tracking engine 130d, a misrecognition engine 130e, and a voice

Google Exhibit 1019
Page 00779

search engine 130f. Furthermore, the conversational language processor 120 may be coupled to one or more data repositories 160 and one or more applications 150 associated with various contexts or domains.

[038] Thus, the system 100 may use the various natural language processing components associated with the conversational language processor 120 in order to engage the user in a cooperative conversation and resolve voice-click device interactions based on the user's intent in initiating the voice-click. More particularly, the intent determination engine 130a may establish meaning for a given multi-modal device interaction based on capabilities of the system 100 as well as the capabilities of any other devices in the multi-device environment. For instance, referring to the example above where the user voice-clicked a particular intersection to determine "What restaurants are around here," the conversational language processor 120 may determine a conversational goal of the voice-click (e.g., "What" may indicate that the utterance relates to a query requesting data retrieval). In addition, the conversational language processor 120 may invoke the context tracking engine 130d to determine context for the voice-click. For example, to determine the voice-click context, the context tracking engine 130d may combine the context associated with the identified point of focus (i.e., the selected intersection) with context associated with the utterance (i.e., restaurants).

[039] As a result, the combined context of the voice-click (which includes both the device interaction and the accompanying utterance) may provide sufficient information for routing a specific query. For example, the query may include various parameters or criteria relating to restaurants and the identified intersection. The conversational language processor 120 may then select a particular device, application, or other component to which the query may be routed for processing. For example, in one implementation, the conversational language processor 120 may evaluate the constellation model 130b, which contains a model of capabilities for each device in the multi-device environment. In one implementation, the constellation model 130b may contain, among other things, knowledge of processing and storage resources available to each of the devices in the environment, as well as the nature and scope of domain agents, context, capabilities, content, services, or other information for each of the devices.

[040] As such, using the constellation model 130b and/or other information, the conversational language processor 120 may determine which device or combination of devices

Google Exhibit 1019
Page 00780

has suitable capabilities that can be invoked to process a given voice-click device interaction. For instance, again referring to the example given above, the conversational language processor 120 may determine that the context of the voice-click relates to an interaction with a navigation device and therefore route the query for processing using a navigation application 150. Results of the query may then be processed (e.g., to weigh the results based on knowledge of the user, such as a preference for vegetarian restaurants) and returned to the user via an output device 180.

[041] According to various aspects of the invention, Figure 2 illustrates a block diagram of an exemplary method for synchronizing multi-modal devices in a natural language voice service environment. As described above, multi-modal device interactions (or "voice-clicks") may generally occur when a user engages in one or more interactions with one or more multi-modal devices while also providing one or more natural language utterances that relate to the interactions with the multi-modal devices. In one implementation, context information relating to the interactions with the multi-modal devices may be combined with context information relating to the natural language utterances to determine an intent of the voice-click (e.g., to initiate a particular action, query, command, task, or other request).

[042] In one implementation, various natural language processing components may be configured to continually listen or otherwise monitor the multi-modal devices to determine when voice-clicks occur. As such, the method illustrated in Figure 2 may be used to calibrate or otherwise configure the components responsible for continually listening or otherwise monitoring the multi-modal devices. For example, in one implementation, the natural language voice service environment may include a plurality of multi-modal devices that provide different capabilities or services, and the user may engage in one or more voice-clicks to request services relating to any of the various devices or capabilities in any given device interaction.

[043] To enable continuous listening for multi-modal device interactions or voice-clicks, each of the plurality of devices in the environment may be configured to receive information relating to a voice-click. Thus, in one implementation, an operation 210 may include establishing device listeners for each of the plurality of devices in the environment. Additionally, operation 210 may be performed in response to one or more new devices being added to the environment. The device listeners established in operation 210 may include any suitable combination of instructions, firmware, or other routines configured to be executed on one or more processing

Google Exhibit 1019
Page 00781

devices or other hardware components. For each device in the environment, the associated device listener may communicate with the device to determine capabilities, features, supported domains, or other information associated with the device. In one implementation, the device listeners may be configured to communicate with the device using the Universal Plug and Play protocol designed for ancillary computer devices. It will be apparent, however, that any suitable mechanism for communicating with the multi-modal devices may be used.

[044] When device listeners have been established for each of the devices in the environment (or when device listeners have been established for new devices added to the environment), the various device listeners may be synchronized in an operation 220. In particular, each of the various devices may have different internal clocks or other timing mechanisms, wherein operation 220 may include synchronizing the various device listeners according to the respective internal clocks or timing mechanisms of the devices. In one implementation, synchronizing the device listeners may include each of the respective device listeners publishing information relating to the internal clock or timing of the associated device.

[045] Thus, when one or more multi-modal interactions or voice-clicks subsequently occur for one or more of the devices, the associated device listener may detect information associated with the voice-click in an operation 230. For example, in one implementation, the various device listeners established in operation 210 may be associated with the voice-click module described above and illustrated in Figure 1. Operation 230 may therefore include one or more of the device listeners or the voice-click module detecting an occurrence of a user interacting with one or more of the devices (e.g., selecting data associated with the device, identifying a point of focus or attention focus associated with the device, or otherwise engaging in one or more interactions or sequences of interactions with the device). Furthermore, an operation 240 may then comprise capturing an utterance from the user that relates to the device interaction detected in operation 230.

[046] For example, a user viewing a web page presented on a display device may see a product name on the web page and desire more information about purchasing the product. The user may select text from the web page that contains the product name (e.g., using a mouse or keyboard to highlight the text), and then initiate a voice-click to ask "Is this available on Amazon.com?" In this example, operation 230 may include a device listener associated with the display device detecting the selection of the text associated with the product name, while

Google Exhibit 1019
Page 00782

operation 240 may include capturing the utterance inquiring about the availability of the product on Amazon.com.

[047] As described above, each device that receives an input from the user may have an internal clock or timing mechanism. In an operation 250, each device may therefore determine when the input was received from a local perspective, and notify the voice-click module that the input was received. In particular, a given voice-click may include at least a natural language utterance received via a voice input device in addition to one or more further interactions with one or more other devices. The utterance may be received prior to, contemporaneously with, or subsequent to the device interactions, whereby operation 250 includes determining the timing of the device interaction for correlation with the associated utterance. In particular, using the device listener signals synchronized as described in reference to operation 220, an operation 260 may include aligning the signals for the device interactions and the utterance. In matching the device interaction and utterance signals, a voice-click input may be generated that includes aligned voice and non-voice components. The voice-click input may then be subject to further natural language processing, as described in greater detail below.

[048] According to various aspects of the invention, Figure 3 illustrates a flow diagram of an exemplary method for processing multi-modal device interactions in a natural language voice service environment. As described above, multi-modal device interactions (or "voice-clicks") may generally occur when a user interacts with one or more multi-modal devices while also providing one or more natural language utterances related to the device interactions. As such, in one implementation, the method illustrated in Figure 3 may be performed when one or more natural language processing components continually listen or otherwise monitor the one or more multi-modal devices to determine when one or more voice-clicks occur.

[049] In one implementation, one or more device interactions may be defined as initiating a voice-click. For example, any given electronic device may generally support various distinguishable interactions, which may result in a given action, command, query, or other request being executed. Thus, in one implementation, any appropriate combination of device interactions that a given electronic device can uniquely recognize or otherwise use to generate a uniquely recognizable signal may be defined as a voice-click, wherein the voice-click may provide a signal indicating that a natural language utterance is to be processed together with context associated with the relevant device interactions. For example, devices that have a four-

Google Exhibit 1019
Page 00783

way or a five-way navigation button may support specific distinguishable interactions, wherein pressing the navigation button in a particular way may cause a specific task or other action to be performed, such as controlling a map display or calculating a route. In another example, a BlackBerry® device having a wheel selector may support interactions such as wheeling a cursor over a particular point of focus or attention focus, pressing the wheel to select specific data or a given application, or various other interactions. Various other device interactions may be used to indicate when a natural language utterance is to be processed together with context associated with the device interactions, without limitation, wherein the specific device interactions may vary in any given implementation. For example, the relevant device interactions may further include one or more of gesturing with a pointing or drawing instrument on a touch-sensitive screen (e.g., drawing an ear-shaped squiggle), a unique interaction method such as a long touch or double-tap, and/or if the system is operating in the continuous listening mode described above, a predefined contextual command word may indicate that a current device context is to be processed with a portion of the voice-based input that follows the contextual command word (e.g., the command word be "OK," "Please," "Computer," or another suitable word, wherein a user may select a specific point on a map and say "Please zoom in," or say "OK read it" when an e-mail is displayed, etc.).

[050] As such, an operation 310 may include processing multi-modal device interactions in a natural language voice service environment to detect an occurrence of one or more device interaction that signal initiation of a voice-click. In particular, the device interactions detected in operation 310 may include any suitable interactions that cause an electronic device to generate a unique, recognizable, or otherwise distinguishable signal that relates to a user selecting data, identifying a point of focus or attention focus, invoking an application or task, or interacting with the device in another way, depending on the specific capabilities of the device.

[051] In addition to the specific signal that the device generates in response to the user interaction, the interaction detected in operation 310 may signal initiation of a voice-click, whereby a prior, contemporaneous, or subsequent natural language voice input will provide further context for interpreting the device interaction detected in operation 310. For example, a natural language processing system may generally be configured to accept a voice input when particular device interactions occur (e.g., pressing a button to turn on a microphone). Thus, in the method illustrated in Figure 3, the device interactions that signal an incoming voice input

Google Exhibit 1019
Page 00784

may further include any suitable interaction or combination of interactions with an electronic device, including interactions associated with a user selecting data, identifying a point of focus or attention focus, invoking an application or task, or interacting with the device in another way, depending on the specific capabilities of the device.

[052] As such, when a voice-click device interaction has been detected in operation 310, a voice-click signal may be generated in an operation 320 to indicate that a natural language voice input should be associated with the interaction detected in operation 320. Subsequently, an operation 330 may include capturing a user utterance to be associated with the interaction detected in operation 310. In one implementation, the interaction detected in operation 310 may signal that a subsequent voice input will be provided, but it will be apparent that in one or more implementations the utterance captured in operation 330 may be provided prior to or contemporaneously with the interaction detected in operation 310 (e.g., a user may provide an utterance such as "Look up this artist on iTunes®" and subsequently voice-click the artist's name on a media player, or the user may provide the utterance while voice-clicking the artist's name, or the user may voice-click the artist's name and then provide the utterance).

[053] When information relating to the voice-click device interactions and associated natural language utterance have been received, an operation 340 may include extracting and combining context information for the device interactions and the associated utterance. In particular, the context information extracted from the voice-click device interactions may include information relating to a segment, item, point of focus, attention focus, or data selected by the user, or to a particular device interaction or sequence of device interactions engaged in by the user. The extracted context for the device interaction may then be combined with context extracted for the natural language utterance captured in operation 330, wherein the combined context information may be used to determine an intent of the voice-click in an operation 350.

[054] For example, in an exemplary voice-click device interaction, a user may be selectively copying a music collection from a media player to a backup storage device. While browsing the music on the media player, the user may encounter a particular song and voice-click the song while saying "Copy this whole album" (e.g., by pressing a particular button on the media player for an extended period of time while highlighting the song). In this example, operation 310 may include detecting the interaction of the extended button press, which causes the voice-click

Google Exhibit 1019
Page 00785

signal to be generated in operation 320. The utterance of "Copy this whole album" may then be captured in operation 330, and context information associated with the voice-click device interaction and the utterance may be combined in operation 340. In particular, the context of the device interaction may include information relating to the selected song, among other things (e.g., the context may further include information contained in metadata associated with the song, such as an ID3 tag for music files). Furthermore, the context of the utterance may include information identifying a copy operation and an album that contains the selected song.

[055] As such, context information relating to voice-click interactions with the multi-modal devices may be combined with context information relating to the natural language utterances, whereby an operation 350 may determine the intent of the voice-click interaction. For instance, referring to the example above, the intent determined in operation 350 may include an intent to copy an album that contains the highlighted song from the media player to the backup storage device. Thus, in response to determining the intent of the voice-click in operation 350, one or more requests may be appropriately routed in an operation 360. In the example being discussed herein, operation 360 may include routing one or more requests to the media player to identify all of the data associated with the album that contains the highlighted song, and one or more requests to any appropriate combination of devices that can manage copying of the identified data from the media player to the backup storage device (e.g., a personal computer that interfaces with both the media player and the storage device).

[056] According to various aspects of the invention, Figure 4 illustrates a flow diagram of an exemplary method for processing multi-modal device interactions to generate a transaction lead or "click-through" in a natural language voice service environment. In particular, the method illustrated in Figure 4 may be used to generate a transaction lead or click-through in combination with one or more actions performed in response to one or more voice-click device interactions being detected.

[057] For example, an operation 410 may include detecting one or more voice-click device interactions received from a user, wherein the voice-click device interactions may comprise any suitable combination of one or more device interactions coupled with one or more related natural language utterances. The user's intent in engaging in the voice-click device interactions may then be determined in an operation 420, and a subsequent operation 430 may include routing one or more requests to one or more processing devices based on the determined

Google Exhibit 1019
Page 00786

intent in order to resolve the voice-click interaction. In one implementation, operations 410, 420, and 430 may be performed in a manner similar to that described above in reference to Figure 2 and Figure 3, whereby signals for the device interactions may be aligned with signals for one or more natural language utterances and context information may be extracted from the signals to determine the intent of the voice-click device interaction.

[058] In addition to routing one or more requests based on the user's intent, the method illustrated in Figure 4 may further comprise generating one or more transaction leads that may result in one or more click-throughs. For example, a click-through may generally refer to an instance of a user clicking or otherwise selecting an electronic advertisement to access one or more services associated with the advertiser. In many electronic systems, click-throughs or click-through rates can provide mechanisms for measuring a user's interactions with an electronic advertisement, which may provide various measurements that advertisers can use to determine an amount to pay an entity that delivers the advertisement to users.

[059] As such, the method illustrated in Figure 4 may generate transaction leads, including advertisements or recommendations, whereby a user's voice-based input combined with a specific device interaction may provide further focus for generating transaction leads. In this manner, the advertisements or recommendations provided to the user may be more relevant to specific information with which the user may be interacting. Furthermore, using natural language cognitive models and shared knowledge relating to a user's preferences may provide further context for targeted transaction leads tailored to the specific user, which may thus be more likely to result in a click-through that can generate payment for a voice services provider.

[060] Thus, in addition to routing one or more requests based on the user's intent in engaging in the voice-click device interaction, an operation 440 may include generating one or more transaction leads based on the determined intent. In particular, based on the combined context of the device interaction and the associated natural language utterance, the transaction leads may be processed in a manner "closer" to the user, in that local voice and non-voice context can be used as state data in any appropriate system that performs targeted advertising. For instance, referring to the example given above where a user selects an intersection displayed on a navigation device while saying "Find restaurants are around here," the transaction leads generated in operation 440 may include one or more advertisements or recommendations for restaurants near the intersection, which may be targeted to the user

Google Exhibit 1019
Page 00787

based on knowledge of the user's short-term and long-term preferences (e.g., a preferred type of restaurant, a preferred price range, etc.).

[061] The transaction leads may then be presented to the user (e.g., as selectable points on a map display). The user's subsequent multi-modal device interactions may then be monitored in an operation 450 to determine whether or when one or more further multi-modal device interactions occur. If no further interactions occur, a determination may be made that the user did not act on the transaction lead and the process may then end. If an additional multi-modal interaction does occur, however, the multi-modal input may be processed in an operation 480 to determine an intent of the input and route one or more requests accordingly. In addition, an operation 460 may include determining whether the multi-modal input relates to the transaction lead generated in operation 440. For example, the user may select one of the advertised or recommended restaurants by providing an utterance, a non-voice device interaction, or a voice-click device interaction requesting further action or information relating to the transaction lead. In such a case, an operation 470 may further include processing a transaction click-through in relation to the transaction lead generated in operation 440, wherein the transaction click-through may be used to determine payment or otherwise generate revenue for a particular entity (e.g., a provider of the voice services or another entity associated with the transaction lead or transaction click-through).

[062] Implementations of the invention may be made in hardware, firmware, software, or various combinations thereof. The invention may also be implemented as instructions stored on a machine-readable medium, which may be read and executed by one or more processors. A machine-readable medium may include various mechanisms for storing or transmitting information in a form readable by a machine (e.g., a computing device). For example, a machine-readable storage medium may include read only memory, random access memory, magnetic disk storage media, optical storage media, flash memory devices, or other storage media, and a machine-readable transmission media may include forms of propagated signals, such as carrier waves, infrared signals, digital signals, or other transmission media. Further, firmware, software, routines, or instructions may be described in the above disclosure in terms of specific exemplary aspects and implementations of the invention, and performing certain actions. However, it will be apparent that such descriptions are merely for convenience, and

Google Exhibit 1019
Page 00788

that such actions in fact result from computing devices, processors, controllers, or other devices executing the firmware, software, routines, or instructions.

[063] Although the descriptions provided herein have generally focused on techniques for processing multi-modal device interactions in a natural language voice services environment, it will be apparent that various further natural language processing capabilities may be used in connection with, in addition to, or in lieu of those described in connection with the specific aspects and implementations discussed herein. For example, in addition to the techniques described in the co-pending U.S. Patent Applications referenced above, the systems and methods described herein may further utilize natural language processing capabilities as described in co-pending U.S. Patent Application 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed August 5, 2005, U.S. Patent Application 11/200,164, entitled "System and Method of Supporting Adaptive Misrecognition in Conversational Speech," filed August 10, 2005, U.S. Patent Application 11/212,693, entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," filed August 29, 2005, U.S. Patent Application 11/580,926, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed October 16, 2006, U.S. Patent Application 11/671,526, entitled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input," filed February 6, 2007, and U.S. Patent Application 11/954,064, entitled "System and Method for Providing a Natural Language Voice User Interface in an Integrated Voice Navigation Services Environment," filed December 11, 2007, the disclosures of which are hereby incorporated by reference in their entirety.

[064] Accordingly, aspects and implementations of the invention may be described herein as including a particular feature, structure, or characteristic, but it will be apparent that every aspect or implementation may or may not necessarily include the particular feature, structure, or characteristic. In addition, when a particular feature, structure, or characteristic has been described in connection with a given aspect or implementation, it will be understood that such feature, structure, or characteristic may be included in connection with other aspects or implementations, whether or not explicitly described. Thus, various changes and modifications may be made to the preceding description without departing from the scope or spirit of the

Google Exhibit 1019
Page 00789

invention, and the specification and drawings should therefore be regarded as exemplary only, with the scope of the invention determined solely by the appended claims.

Google Exhibit 1019
Page 00790

CLAIMS

What is claimed is:

1. A method for processing one or more multi-modal device interactions in a natural language voice services environment that includes one or more electronic devices, comprising:

detecting at least one multi-modal device interaction, wherein the multi-modal device interaction includes a non-voice interaction with at least one of the electronic devices or an application associated with at least one of the electronic devices, and wherein the multi-modal device interaction further includes at least one natural language utterance relating to the non-voice interaction;

extracting context information relating to the multi-modal device interaction, wherein the extracted context information includes context relating to the non-voice interaction, and wherein the extracted context information further include context relating to the natural language utterance;

combining the context relating to the non-voice interaction and the context relating to the natural language utterance;

determining an intent of the multi-modal device interaction based on the combined context relating to the non-voice interaction and the natural language utterance; and

routing at least one request to one or more of the electronic devices based on the determined intent of the multi-modal device interaction.

2. The method of claim 1, wherein at least one of the electronic devices includes an input device configured to receive the natural language utterance.

3. The method of claim 2, further comprising signaling the input device to capture the natural language utterance in response to the non-voice interaction being detected.

4. The method of claim 3, further comprising:

establishing one or more device listeners in the natural language voice services environment, the device listeners configured to detect the non-voice interaction; and

aligning information relating to the non-voice interaction detected by the device listeners and the natural language utterance captured by the input device.

Google Exhibit 1019
Page 00791

5. The method of claim 1, further comprising:

generating at least one transaction lead based on the determined intent of the multi-modal device interaction;

receiving at least one further multi-modal device interaction relating to the generated transaction lead; and

processing a transaction click-through in response to receiving the multi-modal device interaction relating to the generated transaction lead.

6. The method of claim 5, wherein the generated transaction lead includes at least one of an advertisement or a recommendation relating to the determined intent of the multi-modal device interaction.

7. The method of claim 1, wherein the non-voice interaction comprises a selection of a segment, item, data, or application associated with one or more of the electronic devices.

8. The method of claim 1, wherein the non-voice interaction comprises an identification of a point of focus or an attention focus associated with one or more of the electronic devices.

9. The method of claim 1, wherein the non-voice interaction comprises one or more unique and distinguishable interactions associated with one or more of the electronic devices.

10. A system for processing one or more multi-modal device interactions in a natural language voice services environment that includes one or more electronic devices, wherein the system comprises one or more processing devices configured to:

detect at least one multi-modal device interaction, wherein the multi-modal device interaction includes a non-voice interaction with at least one of the electronic devices or an application associated with at least one of the electronic devices, and wherein the multi-modal device interaction further includes at least one natural language utterance relating to the non-voice interaction;

Google Exhibit 1019
Page 00792

extract context information relating to the multi-modal device interaction, wherein the extracted context information includes context relating to the non-voice interaction, and wherein the extracted context information further include context relating to the natural language utterance;

combine the context relating to the non-voice interaction and the context relating to the natural language utterance;

determine an intent of the multi-modal device interaction based on the combined context relating to the non-voice interaction and the natural language utterance; and

route at least one request to one or more of the electronic devices based on the determined intent of the multi-modal device interaction.

11. The system of claim 10, wherein at least one of the electronic devices includes an input device configured to receive the natural language utterance.

12. The system of claim 11, the processing devices further configured to signal the input device to capture the natural language utterance in response to the non-voice interaction being detected.

13. The system of claim 12, the processing devices further configured to:

establish one or more device listeners in the natural language voice services environment, the device listeners configured to detect the non-voice interaction; and

align information relating to the non-voice interaction detected by the device listeners and the natural language utterance captured by the input device.

14. The system of claim 10, the processing devices further configured to:

at least one transaction lead based on the determined intent of the multi-modal device interaction;

receive at least one further multi-modal device interaction relating to the generated transaction lead; and

process a transaction click-through in response to receiving the multi-modal device interaction relating to the generated transaction lead.

Google Exhibit 1019
Page 00793

15. The system of claim 14, wherein the generated transaction lead includes at least one of an advertisement or a recommendation relating to the determined intent of the multi-modal device interaction.

16. The system of claim 10, wherein the non-voice interaction comprises a selection of a segment, item, data, or application associated with one or more of the electronic devices.

17. The system of claim 10, wherein the non-voice interaction comprises an identification of a point of focus or an attention focus associated with one or more of the electronic devices

18. The system of claim 10, wherein the non-voice interaction comprises one or more unique and distinguishable interactions associated with one or more of the electronic devices.

Google Exhibit 1019
Page 00794

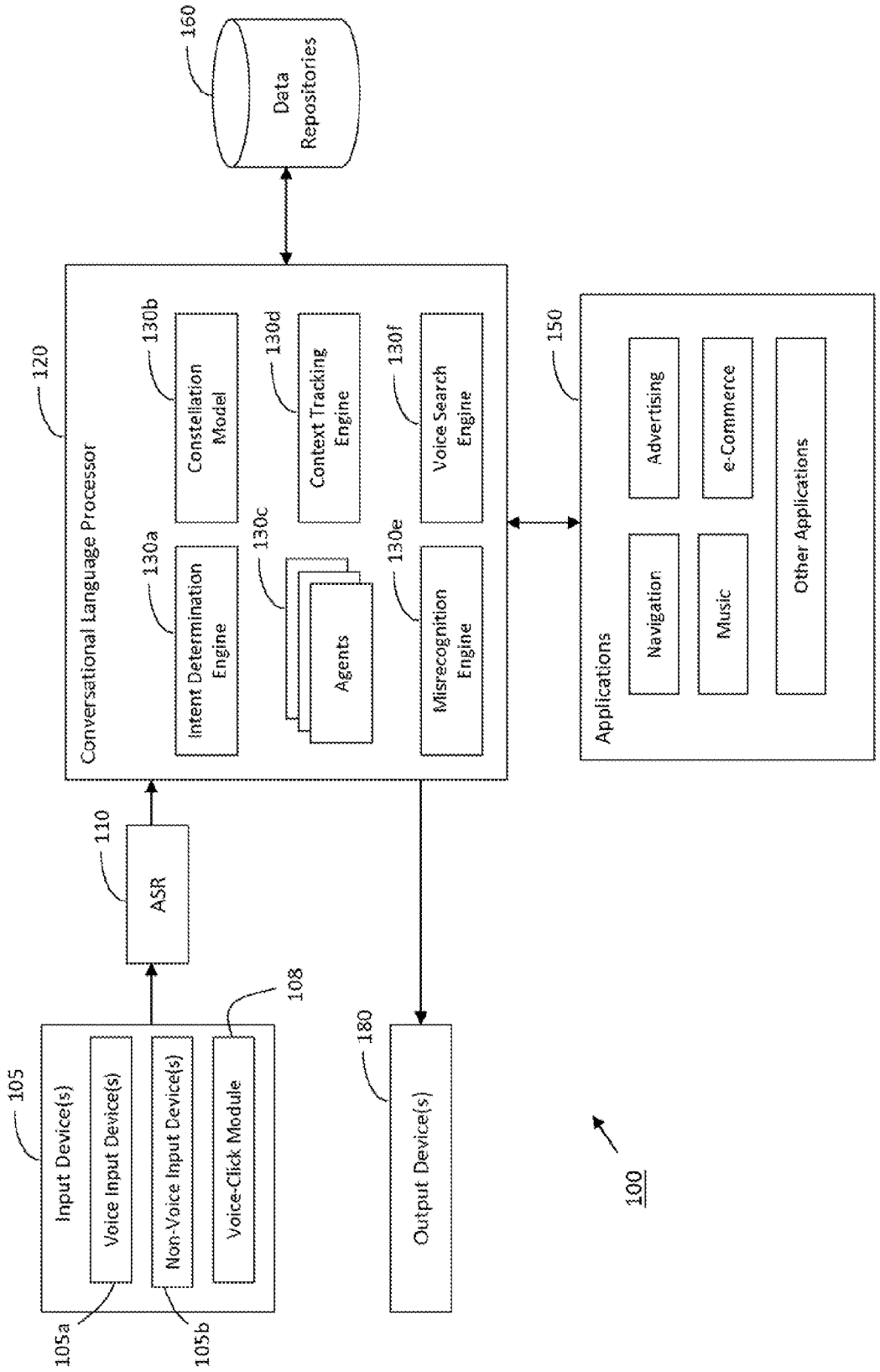

160
Data Repositories
120
Conversational Language Processor
130b
Constellation Model
130d
Context Tracking Engine
130f
Voice Search Engine
130a
Intent Determination Engine
130c
Agents
130e
Misrecognition Engine
150
Applications
Advertising
e-Commerce
Navigation
Music
Other Applications
110
ASR
105
Input Device(s)
Voice Input Device(s)
Non-Voice Input Device(s)
Voice-Click Module
108
180
Output Device(s)
105a
105b
100

Figure 1

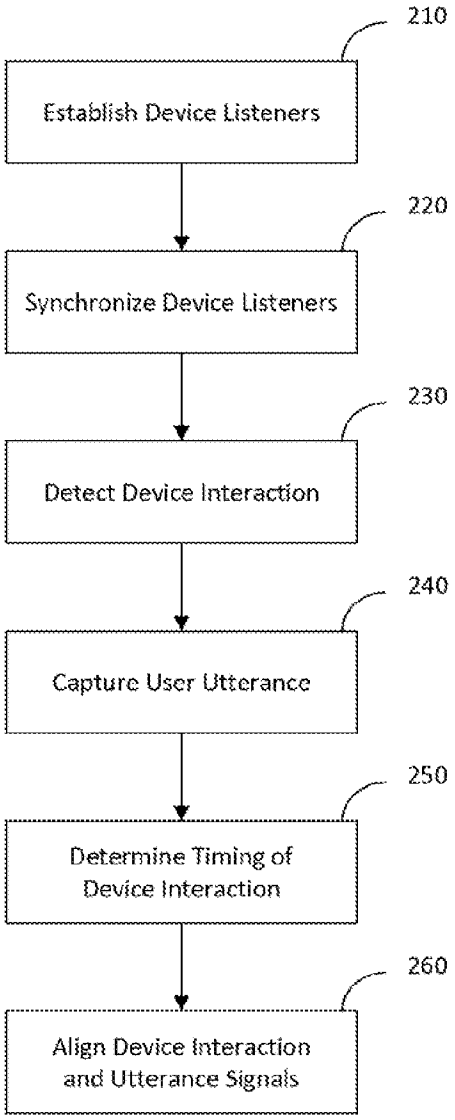

210
Establish Device Listeners
220
Synchronize Device Listeners
230
Detect Device Interaction
240
Capture User Utterance
250
Determine Timing of Device Interaction
260
Align Device Interaction and Utterance Signals

Figure 2

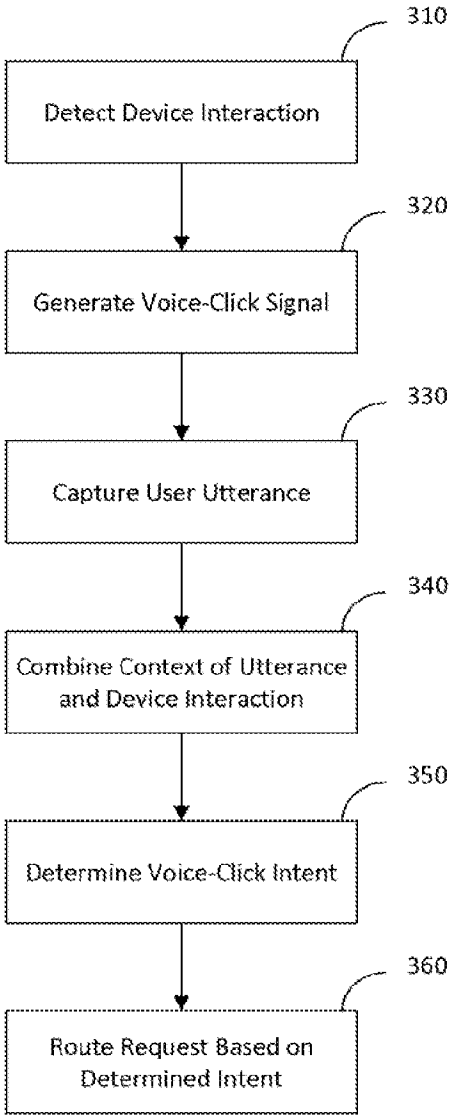

310
Detect Device Interaction
320
Generate Voice-Click Signal
330
Capture User Utterance
340
Combine Context of Utterance and Device Interaction
350
Determine Voice-Click Intent
360
Route Request Based on Determined Intent

Figure 3

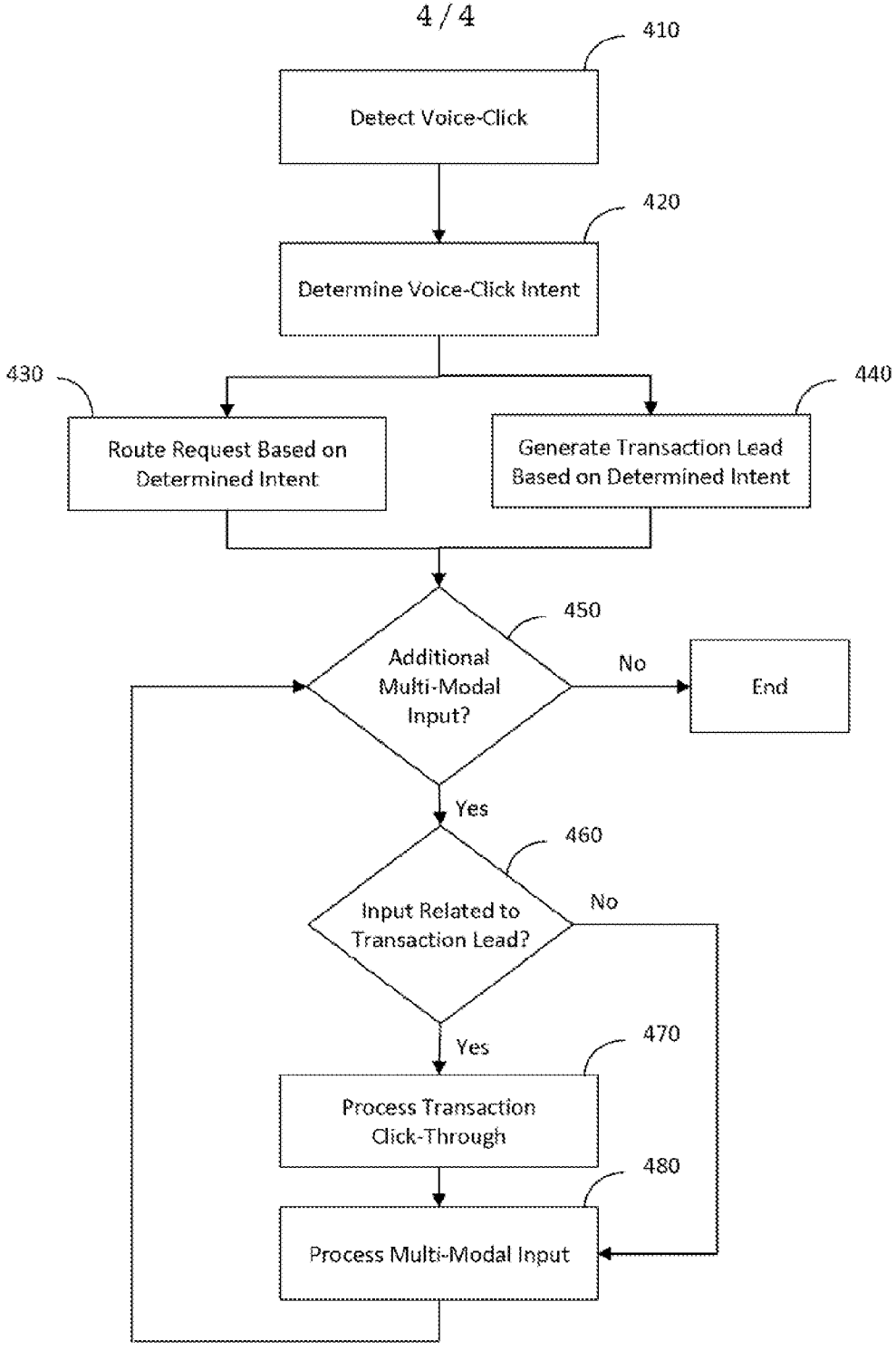

410
Detect Voice-Click
420
Determine Voice-Click Intent
430
Route Request Based on Determined Intent
440
Generate Transaction Lead Based on Determined Intent
450
Additional Multi-Modal Input?
No
End
Yes
460
Input Related to Transaction Lead?
No
Yes
470
Process Transaction Click-Through
480
Process Multi-Modal Input

Figure 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/24895

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G10L 15/00 (2010.01)
USPC - 704/257
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 704/257

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 704/270.1; 704/200; 704/9 (search term limited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic Databases Searched: pubWEST(PGPB,USPT,USOC,EPAB,JPAB); googleScholar
Search Terms Used: natural language, multimodal, speech/voice identification/recognition, voice/keyboard/mouse click, context/intent, marketing/advertising

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2007/0050191 A1 (WEIDER et al.) 01 March 2007 (01.03.2007) Entire document, especially: para [0001], [0016], [0018], [0020], [0031], [0043], [0116], [0133]; Fig. 7, 8	1-2, 7-11 and 16-18 --- 3-6 and 12-15
Y	US 2007/0043574 A1 (COFFMAN et al.) 22 February 2007 (22.02.2007) Entire document, especially: Abstract, para [0154], [0227]	3, 4, 12 and 13
Y	US 2008/0189110 A1 (FREEMAN et al.) 07 August 2008 (07.08.2008) Entire document, especially: para [0006], [0013], [0024], [0028] and Fig. 3	5, 6, 14 and 15
A	US 2008/0235027 A1 (CROSS) 25 September 2008 (25.09.2008)	1-18
A	US 2008/0091406 A1 (BALDWIN et al.) 17 April 2008 (14.04.2008)	1-18

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 March 2010 (28.03.2010)	13 APR 2010
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

Electronic Acknowledgement Receipt

EFS ID:	8971972
Application Number:	12705311
International Application Number:	
Confirmation Number:	8125
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Customer Number:	00909
Filer:	Syed Jafar Ali/Susan Trader
Filer Authorized By:	Syed Jafar Ali
Attorney Docket Number:	090611-0384800
Receipt Date:	06-DEC-2010
Filing Date:	12-FEB-2010
Time Stamp:	14:22:00
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		0384800_IDS.pdf	765051 e1222e6b63443b0297c21773c55f71ca212914e5	yes	14

	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Transmittal Letter		1	2	
	Information Disclosure Statement (IDS) Filed (SB/08)		3	14	
Warnings:					
Information:					
2	Foreign Reference	WO1999046763.pdf	1247910 62eacd3541d472fdc4873f7f386c31e69fd9ea4e	no	28
Warnings:					
Information:					
3	Foreign Reference	WO2000021232.pdf	3963803 203ae5a44439b38f56ae96cdbaa5a64e9e3c570c	no	77
Warnings:					
Information:					
4	Foreign Reference	WO2000046792.pdf	1259047 705348c0646506c1bc55d53c27f1ae60013291b1	no	34
Warnings:					
Information:					
5	Foreign Reference	WO2001078065.pdf	2063508 477bf8d473bf86193d1f7e268aa8aa10625dfa37	no	63
Warnings:					
Information:					
6	Foreign Reference	WO2004072954.pdf	1105346 c4acb3accfd8c87c21d4bf9b2cc1c8dfe629a50c	no	25
Warnings:					
Information:					
7	Foreign Reference	WO2007019318.pdf	3196832 ff1325b971941c7a88c2c7871f5785b0154e9307	no	55
Warnings:					
Information:					
8	Foreign Reference	WO2007021587.pdf	3525793 196b60abf639ab219e2472fadc79b2bf65c0ee3a	no	61
Warnings:					
Information:					

9	Foreign Reference	WO2007027546.pdf	3873530 5cc6e95751b76c1f016b92415f2860dd2cc39298	no	74
Warnings:					
Information:					
10	Foreign Reference	WO2007027989.pdf	1432119 6c33baceb03630bcc7fc958ccf493ab880e7e9e3	no	28
Warnings:					
Information:					
11	Foreign Reference	WO2008098039.pdf	1103202 fcab991006a05016a57e05548f6b00f78732c946	no	24
Warnings:					
Information:					
12	Foreign Reference	WO2008118195.pdf	2058896 635503176929e0d666e3e8683aa40f3a9643658e	no	39
Warnings:					
Information:					
13	Foreign Reference	WO2009075912.pdf	3688456 02c65abf9ea5c876fb8e4a654567b258513f4e25	no	69
Warnings:					
Information:					
14	Foreign Reference	WO2009145796.pdf	1934285 e71f0a22dfc12ce94396146b078264c9e5df68d9	no	39
Warnings:					
Information:					
15	Foreign Reference	WO2010096752.pdf	1502381 f4c719245f739ab6223fbd8f94c7b17cd9acf7b7	no	33
Warnings:					
Information:					
16	NPL Documents	REUTERS.pdf	79394 7f52f8fb293a5d88b570212d1f5399bb61aacdb3	no	1
Warnings:					
Information:					
17	NPL Documents	LIN.pdf	546736 3dce1acf1fbd1bc2a4a0da820e559e124949d760	no	4
Warnings:					
Information:					

18	NPL Documents	KUHN.pdf	438455 5e927839fd4a238b3722e336346ee055e7fcae45	no	5
Warnings:					
Information:					
19	NPL Documents	BELVIN.pdf	468676 93c339f6044e2eb4cff4b1c9087be3a99e3767d4	no	5
Warnings:					
Information:					
20	NPL Documents	LIND.pdf	729916 67eb78f7014eeb103df2382cdcecf0f9807d755a	no	6
Warnings:					
Information:					
21	NPL Documents	ZHAO.pdf	1517578 2808fd320bc985b8850a5a6c3fa515213ffa320c	no	5
Warnings:					
Information:					
22	NPL Documents	CHAI.pdf	1024450 9bf84f1bd4b08cba3cd8e8f3c6d5bb73ea5f9e5b	no	10
Warnings:					
Information:					
23	NPL Documents	CHEYER.pdf	136148 ef2636c9907ae9ab23536f71fecb9453fed36f4a	no	11
Warnings:					
Information:					
24	NPL Documents	ELIO.pdf	934606 5312283c44d9e175a26f467c8014f5de541895b2	no	10
Warnings:					
Information:					
25	NPL Documents	TURUNEN.pdf	179154 01427e429808c0e3724a46047f4d0cfc9abf99d1	no	2
Warnings:					
Information:					
26	NPL Documents	MAO.pdf	87646 07b0d85ca1604fefcd7cd99772e25ff506ff1ef0	no	4
Warnings:					
Information:					

27	NPL Documents	VANHOUCKE.pdf	259233 84ea32d24072d520d3c972cb6af5e8dc4b7835f8	no	4
Warnings:					
Information:					
28	NPL Documents	WENG.pdf	346056 0b8aed950361e7f52843a57fd30ba770384b6363	no	4
Warnings:					
Information:					
29	NPL Documents	ELMELIANI.pdf	296007 f2fb2b6713c7266a4c98b7c64837ebb75106cc03	no	4
Warnings:					
Information:					
30	NPL Documents	ARRINGTON.pdf	627466 b532ea1dde8275e00c1d288d7e7f9a17df60711b	no	4
Warnings:					
Information:					
31	NPL Documents	BAZZI.pdf	361077 67a2e53341c20cd2f0fecfdfc9ca971d47945b8c	no	4
Warnings:					
Information:					
32	NPL Documents	OSHAUGHNESSY.pdf	4563513 93d11484104e08a3f70ac1884e29bccb188a88f0	no	34
Warnings:					
Information:					
			Total Files Size (in bytes):	45316270	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS :	Robert A. KENNEWICK *et al.*	CONFIRMATION NO.:	8125
SERIAL NUMBER :	12/705,311	EXAMINER:	*Unassigned*
FILING DATE :	February 12, 2010	ART UNIT:	2626
FOR :	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE		

INFORMATION DISCLOSURE STATEMENT (IDS)

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, Applicants hereby make of record in the above-identified application the documents listed on the attached Form PTO-1449.

Applicants have listed publication dates on the attached PTO-1449 based on information presently available to the undersigned. However, the listed publication dates should not be construed as an admission that the information was actually published on the date indicated.

Applicants reserve the right to establish the patentability of the claimed invention over any of the information provided herewith, and/or to prove that this information may not be prior art, and/or to prove that this information may not be enabling for the teachings purportedly offered.

This Information Disclosure Statement is being filed within three months of the filing date of a national application other than a continued prosecution application under 37 C.F.R. § 1.53(d), or within three months of the date of entry of the national stage as set forth in 37 C.F.R. § 1.491 in an international application, or before the mailing date of a first Office Action on the merits, or before the mailing date of a first Office Action after the filing of a request for continued examination under 37 C.F.R. § 1.114. No certification or fee is required. 37 C.F.R. § 1.97(b).

It is respectfully requested that the Examiner initial and return a copy of the enclosed PTO-1449.

Respectfully submitted,

PILLSBURY WINTHROP SHAW PITTMAN LLP

By _____
Syed Jafar Ali
Registration No.: 58,780

Customer No.
00909

P. O. Box 10500
McLean, Virginia 22102
Telephone: (703) 770-7900
Telefax: (703) 770-7901

Dated: December 6, 2010

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
12/705,311	02/12/2010	Robert A. Kennewick	090611-0384800	8125

909 7590 12/07/2011

Pillsbury Winthrop Shaw Pittman, LLP (NV)
PO Box 10500
McLean, VA 22102

EXAMINER	
SIRJANI, FARIBA	

ART UNIT	PAPER NUMBER
2626	

NOTIFICATION DATE	DELIVERY MODE
12/07/2011	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket_ip@pillsburylaw.com
margaret.drosos@pillsburylaw.com

Office Action Summary	Application No.	Applicant(s)	
	12/705,311	KENNEWICK ET AL.	
	Examiner	Art Unit	
	FARIBA SIRJANI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *12 February 2010*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

5)☒ Claim(s) *56-85* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *56-85* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *12 February 2010* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

12)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Claims 56-85 are pending. Claims 56 and 71 are independent.

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 56, 58, 71 and 73 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart (U.S. 2002/0069071) in view of Arnold (6,910,003).

Regarding **Claim 56**, **Knockeart** teaches:

56. A mobile system for processing natural language utterances, comprising:

a speech unit connected to a computer device on a vehicle, wherein the speech unit includes an input device that receives a natural language utterance; [Knockeart, Figure 1, "audio input device 110" includes the "speech unit" of the claim. "Audio input device 110 includes a microphone and associated circuitry for processing the audio signal." Co. 3, Lines 29-30. The "controller 130" teaches the "computer device" of the claim. Device 100 also includes a controller 130 that is configured to control the interaction with the user. In addition, controller 130 optionally accepts external condition signals 180 that affect the interaction with the user." Col. 3, Lines 42-45.]

a speech recognition engine that recognizes one or more words or phrases in the natural language utterance; [Knockeart, Figure 1, "audio input device 110." "In some embodiments, audio input device 110 further includes an automatic voice recognizer for interpreting the user's spoken input." Col. 3, Lines 30-33.]

~~a parser~~ that interprets the recognized words or phrases, identifies a telematics context for the natural language utterance from the interpreted words or phrases, and formulates a request in the telematics context from the interpreted words or phrases; and [Knockeart, does not teach a parser. But it does teach that "audio input device 110" interprets the user's input. "In some embodiments, audio input device 110 further includes an automatic voice recognizer for interpreting the user's spoken input." Co. 3, Lines 30-33. Knockeart is directed to telematics and the context of Knockeart is telematics. "As integrated in motor vehicle 200, device 100 can provide an interface to … **telematics** systems such as navigation systems, and telephone systems." Col. 4, Lines 1-4.]

a telematics control unit that interfaces with the computer device on the vehicle, wherein the telematics control unit processes the request in the telematics context. [Knockeart, the controller 130 ("computer device" of the claim) interfaces with telematics that are not shown. The "controller 130" is considered to be the "telematics control unit" of the claim as well. "As integrated in motor vehicle 200, device 100 can provide an interface to a variety of systems. Examples of such systems include **telematics** systems such as navigation systems, and telephone systems. In these examples, user 150 interacts with device 100 to provide inputs to the system, for example, providing inputs to select a destination for a navigation system, or providing inputs to select a telephone number to be dialed by a telephone system. In various embodiments, device 100 can be applied to a wide range of systems in which user 150 needs to provide inputs or feedback to the systems." Col. 4, Lines 1-11.]

Knockeart includes a voice/speech recognizer but does not specifically teach a parser.

Arnold teaches the parser of the claim and also teaches identifying the context of a parsed query. It does not mention telematics but it can identify any context and telematics is just one context which was taught by the primary reference:

a parser that interprets the recognized words or phrases, identifies a ~~telematics~~ context for the natural language utterance from the interpreted words or phrases, and formulates a request in the ~~telematics~~ context from the interpreted words or phrases; [Arnold, Figure 4, "Natural language question parser." "A system, method and article of manufacture are provided for allowing **concept based information searching**. Textual information is collected utilizing a network. The textual information is parsed to create topic specific information packets, which are stored in an information cache. The system parses a user query and compares the parsed user query with the information packets in the information cache to locate matching information packets and displays the matching information packets to a user." Col. 3, Lines 30-33.]

Knockeart and Arnold are both from the art of speech. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the parsing and context determination of Arnold with the speech recognition of Knockeart to expand the capability of Knockeart to distinguishing speech that is directed at telematics and other speech.

Regarding **Claim 58**, Knockeart teaches:

58. The mobile system of claim 56, wherein the telematics control unit processes the request in the telematics context in response to determining that processing the request will not create a hazardous condition for the vehicle. [Knockeart, teaches that telematics requests that are deemed dangerous are rejected: "Device 100 accepts external condition signals 180. Depending on those signals, in some modes of operation device 100 rejects manual inputs. A "switch" 360 in the flowchart in FIG. 3 indicates this selection of modes in which manual input is rejected. For example, in some configurations in which device 100 is integrated in a vehicle (see FIG. 2), manual inputs are rejected when the vehicle speed is too fast for the user to safely use manual inputs and manual input could result in a user being distracted for too long a time at that speed. Note that whether or

not manual inputs are accepted may depend on the nature of those inputs. For example, while it may be safe to provide a simple manual input for which the user does not have to take his eyes off the road such a manual input may not be rejected, but a manual input that requires the driver to select from a list or type an input on a numeric or alphabetic keyboard may be rejected since this would require the driver to potentially unsafely take his eyes off the road.." Col. 4, Lines 47-64.]

Apparatus Claims 56 and 58 and method Claims 71 and 73 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claims 71 and 73 are similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

71. A method for processing natural language utterances, comprising:
receiving a natural language utterance at a speech unit connected to a computer device on a vehicle;
recognizing, at a speech recognition engine connected to the computer device on the vehicle, one or more words or phrases in the natural language utterance;
interpreting, at a parser connected to the computer device on the vehicle, the recognized words or phrases, wherein interpreting the recognized words or phrases include identifying a telematics context for the natural language utterance;
formulating, at the parser, a request in the telematics context from the interpreted words or phrases; and
processing the request in the telematics context with a telematics control unit that interfaces with the computer device on the vehicle.

73. The method of claim 71, wherein the telematics control unit processes the request in the telematics context in response to determining that processing the request will not create a hazardous condition for the vehicle.

3. Claims 57 and 72 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold in view of Sukehiro (2004/0205671).

Regarding **Claim 57**, Knockeart/Arnold teach a telematics control unit, as applied to Claim 56, but Knockeart and Arnold do not teach a multiuser situation with overlapping and interleaved dialogs. **Sukehiro** teaches managing of multiple users in the context of translation requests to translation systems instead of telematics control

unit of the claim. Sukehiro also teaches that the multiple users are served simultaneously (i.e., overlapping/interleaving):

57. The mobile system of claim 56, further comprising

an event manager providing a multi-threaded environment for processing a plurality of requests in a plurality of different contexts across a plurality of overlapping or interleaved dialogs, [Sukehiro, Figure 10, "communication control unit 31," "[0110] The machine translation network system 1 in this embodiment is capable of responding to translation requests from multiple users simultaneously. A single paired machine translation system 14 and HTML parser 17 can operate on a time-sharing basis to respond to multiple translation requests simultaneously, for example, or the system may include multiple pairs of these facilities, which respond to separate translation requests simultaneously. In the latter case, multiple translation requests can be handled simultaneously by loading copies of a machine translation program into the main memories of multiple central processing units (CPUs) with which the server 3 is provided." "[0140] The communication control unit 31 (which includes functions of the HTTP daemon 10, log analyzer 11, communication tools 13a, translation processor 13b, and user registration and authentication unit 13d in FIG. 1) controls communication with the browser and input devices and an external Internet facility (not visible) that stores documents, enabling the retrieval and translation server 23 to retrieve documents from the external Internet facility and supply the retrieved documents or translations thereof to the browser and input devices 24." "[0139] The retrieval and translation server 23 retrieves documents and executes various tasks, including machine translation of the documents. Its component elements include a communication control unit 31, a machine translation unit 32, a dictionary manager 33, a dictionary data base 34, and a terminology incorporator 35."]

wherein the event manager coordinates interaction between the parser and the ~~telematics control unit~~ to enable the ~~telematics control unit~~ to process the request during the overlapping or interleaved dialogs. [Sukehiro, Figure 1 and Figure 10, "communication control unit 31" in Figure 10 which includes 10, 11, 13a and 13b in Figure 1 is shown in Figure 1 to coordinate

between "dictionary manager 33", the "machine translation unit 32" and the "browser and input devices 24". The material input in Sukehiro has to be parsed before it is translated. Sukehiro is translation related instead of telematics related. So, in the combination of prior art pieces, the task done in the context of translation in Sukehiro is combined with the context of telematics from other prior art.]

Sukehiro is directed to natural language processing as are Knockeart and Arnold and it would have been obvious to combine the multi-user capability of Sukehiro that is used in the context of machine translation with the telematics control of Knockeart/Arnold to arrive at the "event manager" of the claim.

Apparatus Claim 57 and method Claim 72 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 72 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

72. The method of claim 71, further comprising providing an event manager having a multi-threaded environment for processing a plurality of requests in a plurality of different contexts across a plurality of overlapping or interleaved dialogs, wherein the event manager coordinates interaction between the parser and the telematics control unit to enable the telematics control unit to process the request during the overlapping or interleaved dialogs.

4. Claims 59 and 74 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold (6,910,003) in view of Lemelson (2002/0022927).

Regarding **Claim 59**, Knockeart teaches the telematics control unit that prevents certain types of manual input in response to detecting a high speed of travel (Col. 4, lines 46-64).

59. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit:

determines whether processing the request in the telematics context will create a hazardous condition for the vehicle; [Knockeart, (Col. 4, lines 46-64)]

…

Knockeart and Arnold do not teach the remainder of the claim but **Lemelson** teaches:

59. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit:

~~determines whether processing the request in the telematics context will create a hazardous condition for the vehicle;~~

provides interactive guidance to resolve the hazardous condition via an output device connected to the vehicle, wherein the telematics control unit provides the interactive guidance to resolve the hazardous condition in response to determining that processing the request in the telematics context will create the hazardous condition; and

[Lemelson, teaches that an accident situation is determined based on position of the vehicle and other data and is communicated to the driver: " GPS satellite (4) ranging signals (6) received (32) on comm1, and DGPS auxiliary range correction signals and pseudolite carrier phase ambiguity resolution signals (8) from a fixed known earth base station (10) received (34) on comm2, at one of a plurality of vehicles/aircraft/automobiles (2) are computer processed (36) to continuously determine the one's kinematic tracking position on a pathway (14) with centimeter accuracy. That GPS-based position is **communicated** with selected other status information **to each other one of the plurality of vehicles** (2), to the one station (10), and/or to one of a plurality of control centers (16), and the one vehicle receives therefrom each of the others' status information and kinematic tracking position. Objects (22) are detected from all directions (300) by multiple supplemental mechanisms, e.g., video (54), radar/lidar (56), laser and optical scanners. Data and information are computer processed and analyzed (50, 52, 200,

452) in neural networks (132, FIGS. 6-8) in the one vehicle to identify, rank, and evaluate collision hazards/objects, an expert operating response to which is determined in a fuzzy logic associative memory (484) which generates control signals which actuate a plurality of control systems of the one vehicle in a coordinated manner to maneuver it laterally and longitudinally to avoid each collision hazard, or, for motor vehicles, when a collision is unavoidable, to minimize injury or damage therefrom. The **operator is warned** by a heads up display and other modes and **may override**. An automotive auto-pilot mode is provided." Abstract. See also paragraph 54.]

processes the request in the telematics context in response to the input device receiving an input that manually overrides the hazardous condition. [Lemelson, " ... The operator is warned by a heads up display and other modes and may **override**. An automotive auto-pilot mode is provided." Abstract. See also paragraph 54.]

Lemelson like the combination of Knockeart/Arnold is directed to telematics and control of vehicles. Accordingly, it would have been obvious to combine Lemelson with the combination of Knockeart/Arnold to arrive at a device with the added safety feature of accident prevention.

Apparatus Claim 59 and method Claim 74 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 74 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

74. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes:

determining whether processing the request in the telematics context will create a hazardous condition for the vehicle;

providing interactive guidance to resolve the hazardous condition via an output device connected to the vehicle, wherein the telematics control unit provides the interactive guidance to resolve the hazardous condition in response to determining that processing the request in the telematics context will create the hazardous condition; and

processing the request in the telematics context in response to the input device receiving an input that manually overrides the hazardous condition.

5. Claims 60-62, 65 and 75-77, 80 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold in view of Obradovichl (7,902,969).

Regarding **Claim 60**, Knockeart teaches that the controller receives sensor data from various parts of the machine [Knockeart, Figure 2, the controller 130 (considered to be both the "computer device" and "telematics control unit" of the claim) is connected to "sensor 280." "In addition to providing an input/output interface to user 150, device 100 receives external condition signals 180 from sensors 280. In this embodiment, sensors 280 include a speed sensor that provides vehicle speed information to device 100." Col. 3, Lines 60-66.] Knockeart and Arnold do not particularly teach that the sensors receive diagnostic information but Obradovichl teaches:

60. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit:

receives diagnostic information from one or more sensors connected to the vehicle, wherein the diagnostic information identifies one or more problems with the vehicle; [Obradovichl, Figure 1, Abstract: "In accordance with the invention, the system provides driving assistance to the user by taking into account the user's physical condition, the vehicle condition and the surrounding conditions.... The vehicle condition concerns the conditions of the brakes, steering, tires, radiator, etc. of the vehicle."]

provides one or more interactive announcements or warnings via an output device connected to the vehicle in response to receiving the diagnostic information that identifies the one or more problems with the vehicle; and [Obradovichl, Figure 20, e.g., "FIG. 20 illustrates the polling routine for monitoring the engine temperature. Instructed by such a routine, processor 103 at step 1603 retrieves temperature data from temperature subsystem 160..... If it is determined that the engine temperature is dangerously high, processor 103 at step 1614 **warns the user**

about the engine overheating condition. At step 1617 processor 103 further instructs the user on how to rectify the condition. For example, instructions such as "Turn on Heater" to assist in cooling down the engine, "Open Windows" to prevent the user from getting too hot inside the car, and "Drive to a Service Station" may be communicated through display 117 and/or user interface 119." Col. 21, Lines 33-47. "User interface 119 comprises conventional audio circuitry including a microphone, voice recognition circuit, voice synthesizer and speaker to allow communications of verbal commands and audio information with system 100." Col. 4, Line 66 to Col. 5, Line 3.]

mediates a solution to resolve the one or more problems with the vehicle. [Obradovichl, "If the temperature data indicates that the engine is dangerously close to overheating or has overheated, processor 103 may instruct the user to "Pull Over" immediately at a safe location." Col. 21, Lines 54-57.]

Obradovichl like the combination of Knockeart/Arnold pertains to control of motor vehicles and uses speech recognition/synthesis in the process. Accordingly, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Obradovich's extensive capabilities with Knockart/Arnold to include the capabilities of Obradovichl in the telematics control of the combination. Note that Obradovich's "user interface 119" in Figure 1 includes speech recognition and synthesis capability but Knockeart is more explicit. (Obradovichl, Col. 4, Line 66 to Col. 5, Line 3: "User interface 119 comprises conventional audio circuitry including a microphone, voice recognition circuit, voice synthesizer and speaker to allow communications of verbal commands and audio information with system 100.")

Regarding **Claim 61**, Obradovichl further teaches:

61. The mobile system of claim 60, wherein the solution to resolve the one or more problems includes one or more of scheduling service for the vehicle, summoning

help to the vehicle over one or more communication links, providing **remedial instructions** that address the one or more problems, ordering one or more parts for the vehicle, or providing an estimated cost for the scheduling the service for the vehicle. [Obradovichl, teaches remedial instructions: "If the temperature data indicates that the engine is dangerously close to overheating or has overheated, processor 103 may instruct the user to "Pull Over" immediately at a safe location." Col. 21, Lines 54-57. "At step 1617 processor 103 further instructs the user on how to rectify the condition. For example, instructions such as "Turn on Heater" to assist in cooling down the engine, "Open Windows" to prevent the user from getting too hot inside the car, and "Drive to a Service Station" may be communicated through display 117 and/or user interface 119." Col. 21, Lines 41-47.]

Same motivation to combine as provided for Claim 60.

Regarding **Claim 62**, Knockart/Arnold, as applied to Claim 56, teach the speech controlled telematics control system but do not teach the specifics of this claim that Obradovichl teaches:

62. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit:

detects an accident situation from information received from one or more sensors connected to the vehicle; [Obradovichl, Figure 23, "FIG. 23 illustrates a safety routine for determining whether the vehicle would clear an object in its path. Such a safety routine is triggered when one such object is detected, e.g., by detection subsystem 130." Col. 23, Lines 48-51. If it does not clear the object, it will collide with it.]

provides one or more interactive announcements or warnings via an output device connected to the vehicle in response to detecting the accident situation, wherein

the telematics control unit provides the interactive announcements or warning to determine one or more of a nature for the accident situation or a condition for a person in the vehicle; and [Obradovichl, Figure 23 "Processor 103 at step 1915 determines whether the lowest vehicle clearance exceeds the height of the object. If so, the routine comes to an end. Otherwise, processor 103 at step 1918 generates a **warning** to be shown on display 117 and/or **announced** through user interface 119, and instructs the user to either negotiate around the object or stop." Col. 23, Lines 57-65.]

reports the nature for the accident situation or the condition for the person in the vehicle to emergency personnel over a communications link. [Obradovichl, The scenario presented in Figure 23 does not expressly include a communication to emergency personnel. However, according to the description of Figure 4, the impaired condition of the driver is communicated to the emergency personnel: "Referring to FIG. 4, recorder subsystem 178 is provided for recording data related to the driver's condition and behavior, the vehicle's condition, and the surrounding conditions..... Processor 2303 may communicate the data to a remote location, via communications subsystem 154. The remote location receives the data from which information is derived for use in reports for fleet management, **police**, insurance, and/or other purposes." Col. 6, Lines 41-59. Police constitutes emergency personnel.]

Same motivation to combine as provided for Claim 60.

Regarding **Claim 65**, Knockart/Arnold, as applied to Claim 56, teach the speech controlled telematics control system but do not teach the specifics of this claim that Obradovichl teaches:

65. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance for performing one or more maneuvers to operate the vehicle via an output device

connected to the vehicle. [Obradovichl, tells the driver how to maneuver around an object on the road: Figure 23, "Processor 103 at step 1915 determines whether the lowest vehicle clearance exceeds the height of the object. If so, the routine comes to an end. Otherwise, processor 103 at step 1918 generates a warning to be shown on display 117 and/or announced through user interface 119, and instructs the user to either negotiate around the object or stop." Col. 23, Lines 57-65.]

Same motivation to combine as provided for Claim 60.

Regarding **Claim 67**, Knockart/Arnold, as applied to Claim 56, teach the speech controlled telematics control system but do not teach the specifics of this claim that Obradovichl teaches:

67. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit:

detects that the vehicle is approaching or has exceeded a travel limit for the vehicle; [Obradovichl, Figure 5 includes a situation where the location of the driver at time t may be determined. From this information it can be detected whether the driver has exceeded a travel limit: "It should be realized that additional data can be analyzed when determining the context of the driver's situation at time t. For example, navigation data from vehicle control subsystem 121 can be used to determine the vehicle's location at time t and indicate whether the driver was on course at that time." Col. 7, Lines 21-26.]

provides interactive guidance for returning to the travel limit via an output device connected to the vehicle in response to detecting that the vehicle is approaching or has exceeded the travel limit; and [Obradovichl, "When the user utilizes navigation system 205 to request instructions for a given destination, the user is elicited for information concerning the destination, any intermediate stops, etc. Such elicitation is realized by posing questions on display 117 and/or by uttering those questions using a synthesized voice through the speaker in user interface 119. The user

then provides verbal responses (or commands) to such questions through the microphone in same. Relying on the aforementioned speech recognition circuitry, navigation system 205 recognizes and registers the responses. Using stored map information, system 205 then provides on display 117 a suggested route leading to the destination. Furthermore, based on the knowledge of the vehicle's instantaneous speeds and directions, system 205 is capable of verbally and visually directing the user to the destination." Col. 5, Lines 39-53. Part of directing the user to the destination is telling the user where he has arrived or has past the destination.]

mediates a negotiation with a person in the vehicle to extend the travel limit. [Obradovichl, Col. 5, Lines 39-53, provided above. The user can provide verbal commands to the navigation system. The user can provide a new destination which would be equivalent to extending the travel limit.]

Same motivation to combine as provided for Claim 60.

Apparatus Claims 60-62, 65, and 67 and method Claims 75-77, 80, and 82 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claims 75-77, 80 and 82 are similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

75. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes:

receiving diagnostic information from one or more sensors connected to the vehicle, wherein the diagnostic information identifies one or more problems with the vehicle;

providing one or more interactive announcements or warnings via an output device connected to the vehicle in response to receiving the diagnostic information that identifies the one or more problems with the vehicle; and

mediating a solution to resolve the one or more problems with the vehicle.

76. The method of claim 75, wherein the solution to resolve the one or more problems includes one or more of scheduling service for the vehicle, summoning help to the vehicle over one or more communication links, providing remedial instructions that address the one or more problems, ordering one or more parts for the vehicle, or providing an estimated cost for the scheduling the service for the vehicle.

77. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes:

detecting an accident situation from information received from one or more sensors connected to the vehicle;

providing one or more interactive announcements or warnings via an output device connected to the vehicle in response to detecting the accident situation, wherein the telematics control unit provides the interactive announcements or warning to determine one or more of a nature for the accident situation or a condition for a person in the vehicle; and

reporting the nature for the accident situation or the condition for the person in the vehicle to emergency personnel over a communications link.

80. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes providing interactive guidance for performing one or more maneuvers to operate the vehicle via an output device connected to the vehicle.

82. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes:

detecting that the vehicle is approaching or has exceeded a travel limit for the vehicle;

providing interactive guidance for returning to the travel limit via an output device connected to the vehicle in response to detecting that the vehicle is approaching or has exceeded the travel limit; and

mediating a negotiation with a person in the vehicle to extend the travel limit.

6. Claims 63 and 78 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart, Arnold and Obradovichl, in view of Odinak (7,406,421).

Regarding **Claim 63**, Obradovichl, as applied to Claim 62, teaches the detection of an emergency situation and establishing communication with the police but is not explicit regarding a voice channel. **Odinak** teaches this feature:

63. The mobile system of claim 62, wherein to further process the request in the telematics control unit, the telematics control unit establishes a voice channel over a communications link to enable communication between the person in the vehicle and the emergency personnel. [Odinak, Figure 1 shows the components and Figure 5 shows that the user can contact a human operator if he wants to (element 220: "Does the user want to contact a human operator?" and "connect user to the human operator 224." "FIG. 1 shows a system 20, a telematics control unit, within a vehicle that performs the present invention. System 20 includes a user interface 22 having a microphone 24 for capturing the user's voice; a display 26; speakers 28; and an interface entry 30 for allowing the user to perform various interactive functions. System 20 also includes a processor 32;

a global positioning system (GPS) 34 for determining precise vehicle locations; and a communications device 36, such as a cellular modem, for transmitting and receiving wireless information. System 20 is also coupled to receive information from other vehicle systems (not shown), such as a speedometer, an engine, or a braking system."]

Odinak is from the art of speech applied in "a vehicle based telematics control unit" as Knockart/Arnold/Obradovich combination is. It would have been obvious to one of ordinary skill at the time of the invention to combine the capability of Odinak regarding live communication with police and emergency personnel in the case of an emergency where time is of the essence.

Apparatus Claim 78 and method Claim 63 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly Claim 78 is similarly rejected under the same rationale as applied above with respect to apparatus Claim 63.

78. The method of claim 77, wherein processing the request in the telematics control unit with the telematics control unit further includes establishing a voice channel over a communications link to enable communication between the person in the vehicle and the emergency personnel.

7. Claims 64 and 79 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold, in view of ObradovichII (6,275,231).

Regarding **Claim 64**, Knockeart and Arnold teach voice recognition in the context of voice controlled devices in telematics but do not address panic and emergency situations. ObradovichII teaches a manual panic/emergency key in a car:

64. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit summons help to the vehicle over one or more communication links in response to determining that the ~~recognized words or~~

~~phrases~~ include at least one predetermined panic word, panic phrase, emergency word, or emergency phrase indicating that a crime is occurring. [ObradovichII, Figure 2 has different keys that may activate different operations: "When emergency key 211g is depressed, processor 103 causes emergency options to be displayed on display 205. These options may include police-type emergencies, medical emergencies, mechanical problems, panic measures, etc. For example, upon selection of the police-type emergencies option, the user may be queried whether he/she is a victim. If the user responds affirmatively, processor 103 immediately establishes a phone or modem connection with a police authority to automatically furnish such preliminary information as the identity of the user, description of the vehicle and emergency contacts, which are pre-recorded, along with the current location of the vehicle identified by navigation system 329. If possible, the user may also add, in the communication, the nature of the emergency and current statuses of the user and any passengers. By automatically communicating at least the preliminary information to the proper authority in case of an emergency, the user can be instantly reached and helped." Col. 10, line 66 to Col. 11, Line 16. See also Col. 11, Lines 17-23 for a panic alarm.]

ObradovichII like the combination of Knockeart/Arnold pertains to control of motor vehicles and uses speech recognition/synthesis in the process. Accordingly, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine ObradovichII's panic and emergency features with Knockart/Arnold to address such situations that are an unfortunate aspect of driving.

Apparatus Claim 64 and method Claim 79 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly Claim 79 is similarly rejected under the same rationale as applied above with respect to apparatus Claim 64.

79. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes summoning help to the vehicle over one or more communication links in

response to determining that the recognized words or phrases include at least one predetermined panic word, panic phrase, emergency word, or emergency phrase indicating that a crime is occurring.

8. Claims 66 and 81 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart, Arnold, and Obradovichl in view of Huang (6,362,748).

Regarding **Claim 66**, Knockeart/Arnold teach the speech controlled telematics system of Claim 56 and Obradovichl teaches:

66. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit:

provides interactive routing information to coordinate a rendezvous with one or more other vehicles in communication with the vehicle; and [Obradovichl teaches a two-way communication with other vehicles regarding road conditions ("It should be noted that the visibility measurement in the subject vehicle may be broadcast to other nearby vehicles using communications subsystem 154 (described below) so that they can benefit from such a measurement. Indeed, the user of the subject vehicle may be a beneficiary of such visibility information received from another vehicle traveling in the same local area." Col. 14, lines 42-52. See also Col. 2, Lines 6-17.) This kind of information includes location information of the vehicle that is reporting the data because it has to tell you where on the road the visibility is going to be bad. Obradovichl thus implicitly teaches reporting the location of a car to other cars.]

~~establishes a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles.~~

Knockeart/Arnold and Obradovichl do not expressly teach calling other cars.

Huang teaches this feature:

establishes a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles. and

[Huang, "The license plate number of a specific vehicle is related to a passcode that is built into a communications reception unit installed in the vehicle. When a driver enters a plate number of the vehicle having the communication reception unit installed, he can communicate with another vehicle. The communication reception unit can receive satellite positioning signals to provide more accurate coordinates to a control center, which can determine the position of the vehicle and send accurate, useful and relevant information to the vehicle." Abstract. "It is thus a primary objective of the invention to provide a system for communicating among vehicles that enables moving vehicles to communicate with each other and can send accurate coordinates to a control center." Col. 1, Lines 56-60. See also Figure 1 and Col. 4, Lines 50-65.]

Huang pertains to communications from a car and during driving and is thus related to telematics of Knockeart/Arnold/ObradovichI combinations. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the cell phone calling capability of Huang with the combination of other references to reduce the danger of diversion of attention during driving.

Apparatus Claim 66 and method Claim 81 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 81 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

81. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes:

providing interactive routing information to coordinate a rendezvous with one or more other vehicles in communication with the vehicle; and

establishing a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles.

9. Claims 68 and 83 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold, in view of ObradovichIII (6,754,485).

Regarding **Claim 68**, Knockeart/Arnold teach the speech enabled telematics system, as applied to Claim 56, but do not teach interactive guidance based on location. **ObradovichIII** teaches an alternative to the OnStar service that uses a car email (C-mail) instead of a cell phone that is used by OnStar and provides interactive information to the driver based on GPS data of the car:

68. The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance relating to one or more tasks for a person in the vehicle based on a location of the vehicle. [ObradovichIII, regarding OnStar, the background states: "The service representative may then provide the user with directions, restaurant recommendations, and/or roadside assistance depending on the vehicle's location." Col. 1, Lines 64-67. Then, the reference teaches that it would use an alternative method of communication: "In accordance with the invention, the communications between a vehicle and remote stations or servers are realized using electronic car-mail, hereinafter referred to as "C-mail."" Col. 4, Lines 33-36. Figure 1 shows the various sensors that can report on the car, Figure 2 shows the communication between the car and a central server, Figure 4 shows a "smart navigator" that can communicate with the user by touch or through a speech interface: "As shown in FIG. 4, the SMART NAVIGATOR page includes displayed options, e.g., NAVIGATION option 419, WEATHER option 421, TRAFFIC option 425 and DIRECTORY option 430. By selecting option 419 by touching same on display 405, followed by selecting ENTER key 435, the user is able to obtain navigation instructions to a given destination. Specifically, the user is elicited for information concerning the destination, any intermediate stops, etc. Such elicitation is realized by posing questions on display 405 and/or by uttering those questions using a synthesized voice through the aforementioned audio output. The user then provides **verbal responses** thereto through the aforementioned audio input. The **speech recognizer** in interface 139 then recognizes and registers the responses." Col. 12, Lines 4-17. The "user interface 139" of Figure 1 has speech capability: "In this instance, interface 139 also includes (1) a conventional **speech**

recognizer and audio input for the user to utter commands and responses to processor 103, and (2) a conventional **speech synthesizer** and audio output for processor 103 to communicate to the user in audio." Col. 11, Lines 10-13. The information provided to the driver is based on the location of the car: "Another message may contain just-in-time map information for navigation, depending on the current location of the automobile. Yet another message may contain advertising information concerning selected entities, e.g., restaurants, gas stations, department stores, etc., also depending on the current location of the automobile. The messages transmitted from the automobile to the remote server may contain, e.g., GPS information identifying the location of the automobile, dynamic data furnished by sensors in the automobile for analysis, an odometer reading and a speedometer reading for records, etc." Abstract.]

Like the combination of Knockeart/Arnold, ObradovichIII is directed to telematics with the use of speech. Thus, it would have been obvious to one of ordinary skill in the art at the time of invention to combined the extended capabilities of ObradovichIII regarding providing guidance to the driver with the simpler telematics system of Knockeart/Arnold to increase the capabilities of this system.

Apparatus Claim 68 and method Claim 83 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 83 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

83. The method of claim 71, wherein processing the request in the telematics control unit with the telematics control unit includes providing interactive guidance relating to one or more tasks for a person in the vehicle based on a location of the vehicle.

10. Claims 69 and 84 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold, in view of Wolfe (6,877,001).

Regarding **Claim 69**, Knockeart/Arnold teach the speech enabled telematics system, as applied to Claim 56, but do not teach the speech recognizer with updatable

dictionary. **Wolfe** teaches "retrieving documents with spoken queries" and keeps its speech recognition dictionary by updating it using the keywords from previously retrieved results (documents) which correspond to previous spoken queries. This way it has a vocabulary that is not static and yet it is not too large:

69. The mobile system of claim 56, wherein the speech recognition engine recognizes the words or phrases in the natural language utterance using a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs. [Wolfe, Figure 1, "Detect keywords and key phrases and construct vocabulary 200," "Dynamic vocabulary and grammar 151" back to the "speech recognition engine 150." "FIG. 2 is a flow diagram of a method for constructing a dynamic speech recognizer vocabulary for an information retrieval system ...." Col. 3, Lines 17-19. "In addition, an active vocabulary and grammar of the speech recognizer or search engine are dynamically updated with key words and key phrases that are automatically extracted from the documents as they are indexed. In other words, information from the document index is fed back into the recognizer or search engine itself." Col. 3, Lines 1-12.]

Wolf is related to the art of speech as is the combination of Knockeart/Arnold and it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the updatable dictionary feature of Arnold with the combination of Knockeart/Arnold to provide a dictionary that is not too massive yet applicable. (See, e.g., Wolfe, paragraph 10).

Apparatus Claim 69 and method Claim 84 are related as apparatus and the method of using same, with each claimed element's function corresponding to the

claimed method step. Accordingly the method Claim 84 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

84. The method of claim 71, wherein the speech recognition engine recognizes the words or phrases in the natural language utterance using a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs.

11. Claims 70 and 85 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold, in view of Wolfe and Shen (7,478,036).

Regarding **Claim 70**, Wolfe teaches that the probabilities of the query words are taken into account in the certainty of the recognition results which in turn generates the search results which yield the dynamic dictionary. [Wolfe, "The invention provides a system and method for retrieving documents from a multi-media database using spoken queries. In addition, the invention makes use of document index information in the speech recognition process, and certainty information about the recognition result while searching for matching documents in the database. The certainty information represents the probabilities of possible query words." Col. 3, Lines 24-59.] Wolfe is not as explicit as **Shen** regarding the use of apriori probabilities that is set forth in Shen in a more straight forward manner:

70. The mobile system of claim 69, wherein the plurality of dictionary and phrase entries are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs. [Shen "C. Filtering out rules may be established based upon prior probabilities. For example, if there is a prior statistic language model derived from standard corpus where we can get Ps(w1 . . . wn), which is the probability of the new extracted word NW=w1 . . . wn, we can calculate Pc(w1 . . . wn) from the current corpus easily. If Pc(w1 . . . wn)/Ps(w1 . . . wn) is large, it means this new word appears relatively more frequently in the current corpus than in standard corpus. So it is a real new word

for this domain. Otherwise, it means this new word combinations are common for standard domain also, so not a new word." Col. 6, Lines 25-35.]

Shen is from the same field of speech as Knockeart/Arnold/Wolfe and it would have been obvious to use the prior probability of the terms in the building of the dictionary considering that this is a usual method in the art and would be use of a known technique to improve similar devices, methods, or products in the same way.

Apparatus Claim 70 and method Claim 85 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 85 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

85. The method of claim 84, wherein the plurality of dictionary and phrase entries are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

McCall (6,591239) teaches a voice controlled device where **manual override** is provided to override the command issued by the voice: [McCall, Abstract, "A voice controlled surgical suite is provided for controlling a plurality of surgical apparatus by a single human. The system includes a voice recognition device adapted to recognize a plurality of predetermined speech commands from a single human and to generate a corresponding set of command output signals….. The surgical apparatus has manual controls to provide redundancy, and to override voice control signals as needed."]

Coffman (7,137,126) teaches simultaneous service to several users using a conversational virtual machine (VM) where the identity of each user is logged (Col. 35,

Lines 4-5) and thus teaches the feature of handling overlapping/interleaved dialogs: [Coffman "Conversational VM can adapt dynamically to the preferences of multiple users and to the active context. It allows multiple users while actively running. In a speech-based system, each command can be used to perform text-independent speaker identification. Any change of user automatically implies the creation of a new active context which pushes the previous context down the context stack, unless the new active context is waived explicitly by the new user or the active application. User changes automatically change the priority along the context stack to first handle a task associated to the active user." Col. 35, Lines 44-54.]

Any inquiry concerning this communication or earlier communications from the examiner should be directed to FARIBA SIRJANI whose telephone number is (571)270-1499. The examiner can normally be reached on Monday through Thursday 9am to 4pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, James Wozniak can be reached on 571-272-7632. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300. Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO

Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FARIBA SIRJANI/
Examiner, Art Unit 2626

/James S. Wozniak/
Primary Examiner, Art Unit 2626

Notice of References Cited	Application/Control No. 12/705,311	Applicant(s)/Patent Under Reexamination KENNEWICK ET AL.	
	Examiner FARIBA SIRJANI	Art Unit 2626	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-7,902,969	03-2011	Obradovich, Michael L.	340/439
*	B	US-6,968,311	11-2005	Knockeart et al.	704/270
*	C	US-6,910,003	06-2005	Arnold et al.	704/4
*	D	US-2004/0205671	10-2004	Sukehiro et al.	715/532
*	E	US-2002/0022927	02-2002	Lemelson et al.	701/301
*	F	US-6,275,231	08-2001	Obradovich, Michael L.	345/156
*	G	US-6,754,485	06-2004	Obradovich et al.	455/414.1
*	H	US-6,362,748	03-2002	Huang, Lin-Hsiang	340/901
*	I	US-2003/0204492	10-2003	Wolf et al.	707/3
*	J	US-7,406,421	07-2008	Odinak et al.	704/275
*	K	US-6,591,239	07-2003	McCall et al.	704/275
*	L	US-7,478,036	01-2009	Shen et al.	704/9
*	M	US-7,137,126	11-2006	Coffman et al.	719/328

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	12705311	KENNEWICK ET AL.
	Examiner	Art Unit
	FARIBA SIRJANI	2626

SEARCHED			
Class	Subclass	Date	Examiner
704	270, 275	11-5-2011	fs

SEARCH NOTES		
Search Notes	Date	Examiner
Inventor search.	11-5-2011	fs
Keyword search of PGPUB, USPAT and Derwent.	11-5-2011	fs
Search advice from M. Opsasnick and J. Wozniak.	11-7-2011	fs

INTERFERENCE SEARCH			
Class	Subclass	Date	Examiner

/FARIBA SIRJANI/ Examiner.Art Unit 2626	

Google Exhibit 1019
Page 00837

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	1A	4,430,669	02/07/1984	Cheung	358	122	
	1B	5,155,743	10/13/1992	Jacobs	375	28	
	1C	5,274,560	12/28/1993	LaRue	364	444	
	1D	5,377,350	12/27/1994	Skinner	395	600	
	1E	5,386,556	01/31/1995	Hedin et al.	395	600	
	1F	5,424,947	06/13/1995	Nagao et al.	364	419.08	
	1G	5,471,318	11/28/1995	Ahuja et al.	358	400	
	1H	5,475,733	12/12/1995	Eisdorfer et al.	379	52	
	1I	5,488,652	01/30/1996	Bielby *et al.*	379	88	
	1J	5,499,289	03/12/1996	Bruno et al.	379	220	
	1K	5,500,920	03/19/1996	Kupiec	395	2.79	
	1L	5,517,560	05/14/1996	Greenspan	379	114	
	1M	5,533,108	07/02/1996	Harris et al.	379	201	
	1N	5,537,436	07/16/1996	Bottoms et al.	375	222	
	1O	5,539,744	07/23/1996	Chu et al.	370	60	
	1P	5,557,667	09/17/1996	Bruno et al.	379	201	
	1Q	5,563,937	10/08/1996	Bruno et al.	379	201	
	1R	5,590,039	12/31/1996	Ikeda et al.	395	759	
	1S	5,617,407	04/01/1997	Bareis	369	275.3	
	1T	5,633,922	05/27/1997	August et al.	379	220	
	1U	5,675,629	10/07/1997	Raffel et al.	379	58	
	1V	5,696,965	12/09/1997	Dedrick	395	610	
	1W	5,708,422	01/13/1998	Blonder et al.	340	825.34	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO
	1X	WO 99/46763	09/16/1999	International				
	1Y	WO 00/21232	04/13/2000	International				
	1Z	WO 00/46792	08/10/2000	International				
	1AA	WO 01/78065	10/18/2001	International				
	1BB	WO 2004/072954	08/26/2001	International				

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

	1CC	REUTERS, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., July 28, 2002, 1 page.
	1DD	LIN, Bor-shen, *et al.*, "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History", ASRU'99, 1999, 4 pages.

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	12/01/2011

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

402527175v1 ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	7	(fuzzy prior aprior) near3 (probability possibility)with (history near6 dialog speech) with dictionary	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 06:18
L2	2	(fuzzy prior aprior) near3 (probability possibility)with (history near6 dialog speech) with dictionary not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 06:18
L3	26	(fuzzy prior aprior) near3 (probability possibility) with dictionary not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 06:19
L4	2	(fuzzy prior aprior) near3 (probability possibility) and glushnev.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 06:40
L5	22	(fuzzy prior aprior) near3 (probability possibility) and 704/10.ccls.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 06:41
L6	25506	((multiple several many) near4 (user speaker)) and ((voice speech) near5 (recogni$4))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 07:32
L7	4513	((multiple several many) near4 (user speaker dialog)) and ((voice speech) near5 (recogni$4)) and "704".clas.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 07:32
L8	48	((multiple several many) near4 (user speaker dialog)) and ((voice speech) near5 (recogni$4)) and "704".clas. and ((overlap$4 simultaneous$4) near4 (dialog))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 07:33
L9	48	((multiple several many) near3 (user speaker dialog)) and ((voice speech) near3 (recogni$4)) and "704".clas. and ((overlap$4 simultaneous$4) near4 (dialog))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 07:39
L10	47	((multiple several many) near3 (user speaker dialog)) and ((voice speech) near3 (recogni$4)) and "704".clas. and ((overlap$4 simultaneous$4) near4 (dialog)) not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 07:39
L11	51	((multiple several many) adj3 (user speaker dialog)) and "704".clas. and ((overlap$4 simultaneous$4) near4 (dialog)) not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/14 08:01
S1	2758	704/270.ccls.	US-PGPUB;	OR	ON	2011/11/05 13:25

			USPAT; DERWENT			
S2	35	Kennewick.in. near robert	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:25
S3	5	Kennewick.in. near michael	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:25
S4	17	Kennewick.in. near richard	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S5	91	locke.in. near david	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S6	19	freeman.in. near tom	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S7	12	elston.in. near stephen	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S8	1	S2 and S3 and S4 and S5 and S6 and S7	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S9	2	"7693720".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:27
S10	9	("7693720").URPN.	USPAT	OR	ON	2011/11/05 13:27
S11	9	("7693720").URPN.	USPAT	OR	ON	2011/11/05 13:27
S12	187	("20010041980" | "20020049805" | "20020065568" | "20020082911" | "20020120609" | "20020124050" | "20020143535" | "20020188602" | "20020198714" | "20030014261" | "20030112267" | "20030182132" | "20040025115" | "20040044516" | "20040166832" | "20040199375" | "20050021334" | "20050021826" | "20050033574" | "20050114116" | "20050246174" | "20070033005" | "20070038436" | "20070050191" | "20080091406" | "20080103761" | "20080115163" | "20080133215" | "20090117885" | "20090144271" | "4430669" | "5155743" | "5274560" | "5377350" | "5386556" | "5424947" | "5471318" | "5475733" | "5499289" | "5500920" | "5517560" | "5533108" | "5537436" | "5539744" | "5557667" | "5563937" | "5590039" | "5617407" | "5633922" | "5675629" | "5696965" |	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/05 13:27

		"5708422" | "5721938" | "5722084" | "5742763" | "5752052" | "5754784" | "5761631" | "5774859" | "5794050" | "5797112" | "5799276" | "5802510" | "5832221" | "5878385" | "5878386" | "5892813" | "5895466" | "5902347" | "5911120" | "5918222" | "5926784" | "5933822" | "5953393" | "5963894" | "5963940" | "5987404" | "5991721" | "5995119" | "6009382" | "6014559" | "6021384" | "6044347" | "6049602" | "6058187" | "6078886" | "6081774" | "6101241" | "6119087" | "6134235" | "6144667" | "6160883" | "6173279" | "6175858" | "6185535" | "6192110" | "6192338" | "6195634" | "6195651" | "6208972" | "6219346" | "6219643" | "6233556" | "6233559" | "6246981" | "6272455" | "6292767" | "6314402" | "6366886" | "6381535" | "6385646" | "6393428" | "6404878" | "6408272" | "6411810").PN. OR ("6415257" | "6418210" | "6420975" | "6430285" | "6434523" | "6434524" | "6442522" | "6446114" | "6453153" | "6453292" | "6456711" | "6466654" | "6466899" | "6498797" | "6499013" | "6501833" | "6501834" | "6513006" | "6523061" | "6532444" | "6539348" | "6553372" | "6556970" | "6556973" | "6560576" | "6567778" | "6567797" | "6570555" | "6570964" | "6574597" | "6574624" | "6594257" | "6598018" | "6604077" | "6611692" | "6614773" | "6615172" | "6629066" | "6631346" | "6643620" | "6650747" | "6691151" | "6721001" | "6721706" | "6735592" | "6741931" | "6742021" | "6757718" | "6795808" | "6833848" | "6865481" | "6877134" | "6901366" | "6937977" | "6944594" | "6973387" | "6980092" | "6990513" | "6996531" | "7020609" | "7027975" | "7062488" | "7092928" | "7127400" | "7137126" | "7146319" | "7289606" | "7376645" | "7398209" | "7424431" | "7461059" | "7493559").PN.				
S13	187	("20010041980" | "20020049805" | "20020065568" | "20020082911" | "20020120609" | "20020124050" | "20020143535" | "20020188602" | "20020198714" | "20030014261" | "20030112267" | "20030182132" | "20040025115" | "20040044516" | "20040166832" | "20040199375" | "20050021334" | "20050021826" | "20050033574" | "20050114116" | "20050246174" | "20070033005" | "20070038436" | "20070050191" | "20080091406" | "20080103761" | "20080115163" | "20080133215" | "20090117885" | "20090144271" | "4430669" | "5155743" | "5274560" | "5377350" | "5386556" | "5424947" | "5471318" | "5475733" | "5499289" |	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/05 13:27

		"5500920" | "5517560" | "5533108" | "5537436" | "5539744" | "5557667" | "5563937" | "5590039" | "5617407" | "5633922" | "5675629" | "5696965" | "5708422" | "5721938" | "5722084" | "5742763" | "5752052" | "5754784" | "5761631" | "5774859" | "5794050" | "5797112" | "5799276" | "5802510" | "5832221" | "5878385" | "5878386" | "5892813" | "5895466" | "5902347" | "5911120" | "5918222" | "5926784" | "5933822" | "5953393" | "5963894" | "5963940" | "5987404" | "5991721" | "5995119" | "6009382" | "6014559" | "6021384" | "6044347" | "6049602" | "6058187" | "6078886" | "6081774" | "6101241" | "6119087" | "6134235" | "6144667" | "6160883" | "6173279" | "6175858" | "6185535" | "6192110" | "6192338" | "6195634" | "6195651" | "6208972" | "6219346" | "6219643" | "6233556" | "6233559" | "6246981" | "6272455" | "6292767" | "6314402" | "6366886" | "6381535" | "6385646" | "6393428" | "6404878" | "6408272" | "6411810").PN. OR ("6415257" | "6418210" | "6420975" | "6430285" | "6434523" | "6434524" | "6442522" | "6446114" | "6453153" | "6453292" | "6456711" | "6466654" | "6466899" | "6498797" | "6499013" | "6501833" | "6501834" | "6513006" | "6523061" | "6532444" | "6539348" | "6553372" | "6556970" | "6556973" | "6560576" | "6567778" | "6567797" | "6570555" | "6570964" | "6574597" | "6574624" | "6594257" | "6598018" | "6604077" | "6611692" | "6614773" | "6615172" | "6629066" | "6631346" | "6643620" | "6650747" | "6691151" | "6721001" | "6721706" | "6735592" | "6741931" | "6742021" | "6757718" | "6795808" | "6833848" | "6865481" | "6877134" | "6901366" | "6937977" | "6944594" | "6973387" | "6980092" | "6990513" | "6996531" | "7020609" | "7027975" | "7062488" | "7092928" | "7127400" | "7137126" | "7146319" | "7289606" | "7376645" | "7398209" | "7424431" | "7461059" | "7493559").PN.				
S14	196	S11 or S12	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:28
S15	3	"6615172".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 06:00
S16	117	"s196" and pars$4	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 06:02
S17	82	"s196" and pars$4 and (GPS telematics	US-	OR	ON	2011/11/07

		navigat$4)	PGPUB; USPAT; DERWENT			06:03
S18	2	"7693720".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 06:08
S19	1	S18 and pars$4 and (GPS telematics navigat$4)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 06:08
S20	23	"10/452147"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 06:25
S21	2	"20100145700".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 07:12
S22	2	"7693720".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 09:13
S23	2	"6937977".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 09:41
S24	4263	telematics.ti,abst,bsum,clm.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:49
S25	140	telematics.ti,abst,bsum,clm. and "704".clas.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:49
S26	1732	telematics.ti,abst,bsum,clm. and (command query)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:50
S27	102	telematics.ti,abst,bsum,clm. and (command query) and "704".clas.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:50
S28	28	telematics.ti,abst,bsum,clm. and (command query) and "704".clas. and safety	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:51
S29	71	telematics.ti,abst,bsum,clm. and (command query) and "704".clas. and (safety hazard$4 emergency panic)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:52
S30	72	telematics.ti,abst,bsum,clm. and (command query warn$4) and "704".clas. and (safety hazard$4 emergency panic)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:53
S31	73	telematics.ti,abst,bsum,clm. and (command query warn$4) and "704".clas. and (safety hazard$4 emergency panic accident)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:53

S32	76	telematics.ti,abst,bsum,clm. and (command query warn$4 sensor) and "704".clas. and (safety hazard$4 emergency panic accident)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:54
S33	77	telematics.ti,abst,bsum,clm. and (command query warn$4 sensor) and "704".clas. and (safety hazard$4 emergency panic accident travel)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:54
S34	6	telematics.ti,abst,bsum,clm. and (command query warn$4 sensor) and "704".clas. and (safety hazard$4 emergency panic accident travel) and ((dialogue interactive convers$4) near5 (history record))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:55
S35	71	telematics.ti,abst,bsum,clm. and (command query warn$4 sensor) and "704".clas. and (safety hazard$4 emergency panic accident travel) not kennewick	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:56
S36	71	telematics.ti,abst,bsum,clm. and (command query warn$4 sensor) and "704".clas. and (safety hazard$4 emergency panic accident travel) not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 10:56
S37	134	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 11:18
S38	63	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not S36	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 11:19
S39	52	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not S36 not rhoads.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 11:19
S40	105	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not rhoads.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 12:15
S41	105	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not rhoads.in. and ay>="2002"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 12:16
S42	105	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not rhoads.in. and fy>="2002"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 12:17
S43	105	telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not rhoads.in. and fy<="2002"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 12:17
S44	105	fy<="2002" and telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not rhoads.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 12:17
S45	7	@py<="2002" and telematics.ti,abst,bsum,clm. and "704".clas. not kennewick.in. not	US-PGPUB; USPAT;	OR	ON	2011/11/07 12:32

		rhoads.in.	DERWENT			
S46	275	@py<="2002" and telematics.ti,abst,bsum,clm. not kennewick.in. not rhoads.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/07 14:15
S47	2	09/918,360	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 08:37
S48	275	@py<="2002" and telematics.ti,abst,bsum,clm. not kennewick.in. not rhoads.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 09:20
S49	16	S48 and ((speech voice) near3 (recogni$4)) and mobile	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 09:20
S50	10	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and (telematic$4 onstar on-star) and @py<="2007"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 09:51
S51	846	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2007"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 09:51
S52	39	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and (telematic$4 onstar on-star)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 09:51
S53	10	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and (telematic$4 onstar on-star) and @py<="2007"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 13:46
S54	39	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and (telematic$4 onstar on-star)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 13:46
S55	29	S54 not S53	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 13:46
S56	839	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2007" not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 13:49
S57	541	(car navig$4 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2005" not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 13:49
S58	608	(car navig$5 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2005" not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 14:15
S59	596	(car navig$5 vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2005" not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 14:22

		not parry.in.				
S60	534	(car navigation navigator vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2005" not kennewick.in. not parry.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 14:36
S61	118	(car navigation navigator vehicle gps) and ((speech voice) near3 (recogni$4)) and (NLU (natural near3 language)) and (pars$4) and @py<="2005" not kennewick.in. not parry.in. and (mobile near3 device)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 14:42
S62	2	"7275049".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/08 14:59
S63	15	("7275049").URPN.	USPAT	OR	ON	2011/11/08 14:59
S64	15	("7275049").URPN.	USPAT	OR	ON	2011/11/08 14:59
S65	10	("20030204492" | "20040162724" | "20040193420" | "5003506" | "5799276" | "6173251" | "6327343" | "6510417" | "6901364" | "7103563").PN.	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/08 14:59
S66	10	("20030204492" | "20040162724" | "20040193420" | "5003506" | "5799276" | "6173251" | "6327343" | "6510417" | "6901364" | "7103563").PN.	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/08 14:59
S67	25	S64 or S65	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/08 14:59
S68	15	("7275049").URPN.	USPAT	OR	ON	2011/11/09 08:57
S69	10	("20030204492" | "20040162724" | "20040193420" | "5003506" | "5799276" | "6173251" | "6327343" | "6510417" | "6901364" | "7103563").PN.	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/09 08:57
S70	24	S68 or S69 not kennewick.in.	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/09 08:57
S71	3	"6901364".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/09 14:16
S72	3	"6910003".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/09 14:21
S73	41	(US-20020188392-$ or US-20020175647-$ or US-20020173889-$ or US-20020164973-$ or US-20020151297-$ or US-20020140545-$ or US-20020140215-$ or US-20020120374-$ or US-20020115436-$	US-PGPUB; USPAT	OR	ON	2011/11/10 07:29

		or US-20020091706-$ or US-20020091486-$ or US-20020086680-$ or US-20020075168-$ or US-20020069312-$ or US-20020069071-$ or US-20020044065-$ or US-20020032510-$ or US-20080033770-$ or US-20050262065-$ or US-20050256889-$ or US-20040230494-$ or US-20040215449-$ or US-20040205671-$ or US-20040199375-$ or US-20040168119-$ or US-20030204492-$).did. or (US-7219063-$ or US-6968311-$ or US-6501387-$ or US-RE39942-$ or US-6981040-$ or US-6977992-$ or US-6976053-$ or US-6970935-$ or US-6910003-$ or US-6885734-$ or US-6801604-$ or US-6792082-$ or US-6745161-$ or US-6901364-$ or US-6173251-$).did.				
S74	17	S73 and ((multiple several many) near3 (user input entry dialog))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 07:30
S75	12	S73 and ((multiple several many) near3 (user input entry dialog)) and pars$4	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 07:58
S76	2	"6185535".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 09:50
S77	0	704/275.clas.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 10:27
S78	2577	704/275.ccls.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 10:28
S79	1168	704/275.ccls. and @py<="2005"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 10:28
S80	813	704/275.ccls. and @py<="2005"	USPAT	OR	ON	2011/11/10 10:28
S81	5	704/275.ccls. and @py<="2005" and (telematic telematics)	USPAT	OR	ON	2011/11/10 10:28
S82	10	704/275.ccls. and @py<="2005" and (telematic telematics)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 12:36
S83	2	704/275.ccls. and (telematic telematics) and (manual near3 override)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 12:37
S84	11	704/275.ccls. and (manual near3 override)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 12:37
S85	9	704/275.ccls. and (manual near3	US-	OR	ON	2011/11/10

		override) not kennewick.in.	PGPUB; USPAT; DERWENT			12:38
S86	818	704/275.ccls. and (interactive) not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 12:53
S87	316	704/275.ccls. and (interactive) and (car vehicle auto$4) not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 12:54
S88	163	704/275.ccls. and (interactive) and (car vehicle auto$4).ti,abst,bsum,clm. not kennewick.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 13:48
S89	2	"6185535".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:22
S90	2	"6980092".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:24
S91	2	"20020151297".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:27
S92	2	"6466232".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:33
S93	2	"6108532".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:33
S94	2	"6718187".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:34
S95	2	"6546257".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:35
S96	2	"6405106".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:35
S97	2	"6389278".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/10 14:37
S98	2	"20010023397".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 07:16
S99	2	"20060122834".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 07:18

S100	2	"5797116".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 07:38
S101	5	"7302383".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 07:44
S102	5	("20010011217" | "20020165862" | "20030125955" | "6615172" | "6904428").PN.	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/11 07:45
S103	6	("7302383").URPN.	USPAT	OR	ON	2011/11/11 08:14
S104	614	calcar	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:14
S105	147	calcar and (car vehicle automobile)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:15
S106	130	calcar and (car vehicle automobile) not polymer	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:15
S107	130	calcar and (car vehicle automobile) not polymer not calcar.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:33
S108	119	calcar and (car vehicle automobile) not polymer and obradovich.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:33
S109	142	(car vehicle automobile) not polymer and obradovich.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:33
S110	178	obradovich.in.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:33
S111	135	obradovich.in. near michael	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:34
S112	39	obradovich.in. near michael and accident	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 08:34
S113	100	(obradovich.in. near michael) and (hazard$4 danger$4 accident$4)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 10:18
S114	49	(obradovich.in. near michael) and ((hazard$4 danger$4 accident$4) near3 condition)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 10:19

S115	1	(obradovich.in. near michael) and ((danger$4 accident$4) near3 condition)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 10:34
S116	10	"7,902,969".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 10:51
S117	10	"7902969".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 10:51
S118	10	"7902969".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 10:51
S119	39	obradovich.in. near michael and accident	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 15:43
S120	480134	obradovich.in. near michael and (panic emergency crime)]	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 18:25
S121	125	obradovich.in. near michael and (panic emergency crime)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 18:25
S122	44	obradovich.in. near michael and (panic crime)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 18:26
S123	29	obradovich.in. near michael and (panic crime) not (crime adj area)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 18:26
S124	27	obradovich.in. near michael and (panic crime) not (crime adj (rate area))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/11 18:27
S125	9	obradovich.in. near michael and ((interactive near5 guidance) (negotiat$4))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 07:03
S126	0	obradovich.in. near michael and ((interactive near5 guidance) (negotiat$4)) not (negotiat$4 adj3 corner)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 07:04
S127	0	obradovich.in. near michael and (interactive near5 guidance)	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 07:04
S128	20	obradovich.in. near michael and onstar	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 07:04
S129	9	obradovich.in. near michael and ((call$4 communicat$4) with (other nearby) with (car vehicle automobile))	US-PGPUB; USPAT;	OR	ON	2011/11/12 09:10

			DERWENT			
S130	2290	((call$4 communicat$4) with (other nearby) with (car vehicle automobile))	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 09:12
S131	872	((call$4 communicat$4) with (other nearby) with (car vehicle automobile)) and @py<="2005"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 09:12
S132	467	((call$4 communicat$4) with (other nearby) with (car vehicle automobile)) and @py<="2002"	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 09:12
S133	75	((call$4 communicat$4) with (other nearby) with (car vehicle automobile)) and @py<="2002" and accident	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 10:12
S134	88	(US-20020188392-$ or US-20020175647-$ or US-20020173889-$ or US-20020164973-$ or US-20020151297-$ or US-20020140545-$ or US-20020140215-$ or US-20020120374-$ or US-20020115436-$ or US-20020091706-$ or US-20020091486-$ or US-20020086680-$ or US-20020075168-$ or US-20020069312-$ or US-20020069071-$ or US-20020044065-$ or US-20020032510-$ or US-20080033770-$ or US-20050262065-$ or US-20050256889-$ or US-20040230494-$ or US-20040215449-$ or US-20040205671-$ or US-20040199375-$ or US-20040168119-$ or US-20030204492-$).did. or (US-20100312566-$ or US-20090313026-$ or US-20090216540-$ or US-20060190269-$ or US-20050257174-$ or US-20040215464-$ or US-20030083884-$ or US-20040193426-$ or US-20020022927-$).did. or (US-7219063-$ or US-6968311-$ or US-6501387-$ or US-RE39942-$ or US-6981040-$ or US-6977992-$ or US-6976053-$ or US-6970935-$ or US-6910003-$ or US-6885734-$ or US-6801604-$ or US-6792082-$ or US-6745161-$ or US-6901364-$ or US-6173251-$ or US-6839670-$ or US-7406421-$ or US-7920682-$ or US-7801731-$ or US-7769143-$ or US-7426467-$ or US-6411934-$ or US-5812977-$ or US-4797924-$ or US-6108532-$ or US-6405106-$).did. or (US-7902969-$ or US-7802198-$ or US-7748021-$ or US-7741959-$ or US-7739039-$ or US-7702455-$ or US-7656310-$ or US-7650234-$ or US-7593812-$ or US-7561065-$ or US-7535344-$ or US-7522992-$ or US-7499778-$ or US-7461561-$ or US-7475057-$ or US-7417531-$ or US-	US-PGPUB; USPAT	OR	ON	2011/11/12 11:54

		7409221-$ or US-7343165-$ or US-7319848-$ or US-7251789-$ or US-6282464-$ or US-6577928-$ or US-7171291-$ or US-6754485-$ or US-6405132-$ or US-6401036-$ or US-6370475-$).did.				
S135	2	S134 and wolfe	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 11:54
S136	2	"6877001".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/12 11:57

11/14/2011 8:44:03 AM
C:\Users\fsirjani\Documents\EAST\Workspaces\1s.wsp

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	12705311	KENNEWICK ET AL.
	Examiner	Art Unit
	FARIBA SIRJANI	2626

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	11/05/2011								
	56	✓								
	57	✓								
	58	✓								
	59	✓								
	60	✓								
	61	✓								
	62	✓								
	63	✓								
	64	✓								
	65	✓								
	66	✓								
	67	✓								
	68	✓								
	69	✓								
	70	✓								
	71	✓								
	72	✓								
	73	✓								
	74	✓								
	75	✓								
	76	✓								
	77	✓								
	78	✓								
	79	✓								
	80	✓								
	81	✓								
	82	✓								
	83	✓								
	84	✓								
	85	✓								

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE PATENT APPLICATION OF:	Robert A. KENNEWICK *et al.*
SERIAL NO.:	12/705,311
FILING DATE:	February 12, 2012
ATTORNEY DOCKET NO.:	090611-0384800
CONFIRMATION NO.:	8125
EXAMINER:	Fariba Sirjani
ART UNIT:	2626
FOR:	Mobile Systems and Methods for Responding to Natural Language Speech Utterance

RESPONSE TO NON-FINAL OFFICE ACTION mailed 07 December 2011

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA. 22313-1450

Dear Madam:

In response to the Office Action mailed **December 7, 2011**, the date for responding being March 7, 2012, with a two-month extension of time filed herewith, please amend the above- identified application as follows:

Amendments to the Claims begin on page **2** of this paper.

Remarks begin on page **13** of this paper.

This Response is being filed with a Petition for Extension of Time (two months). However, if additional extensions of time are necessary to prevent abandonment of this application, then such extensions of time are hereby petitioned for under 37 C.F.R. § 1.136(a), and any fees required therefore (including fees for net addition of claims) are hereby authorized to be charged to Deposit Account No. 033975 (**Ref. No. 090611-0384800**).

This listing of claims will replace all prior versions and listings of claims in the Application.

LISTING OF CLAIMS:

1-55. (**Cancelled**)

56. (**Currently Amended**) A mobile system for processing natural language utterances, comprising:

one or more processors configured to:

receive a natural language utterance,

~~a speech unit connected to a computer device on a vehicle, wherein the speech unit includes an input device that receives a natural language utterance;~~

execute a speech recognition engine that recognizes one or more words or phrases in the natural language utterance, [[;]]

execute a parser that ~~interprets the recognized words or phrases,~~ identifies a telematics domain from the recognized one or more words or phrases of the natural language utterance ~~context for the natural language utterance from the interpreted words or phrases,~~ and formulates a request based on ~~in~~ the telematics domain ~~context from the interpreted words or phrases~~; and

communicate the request to a telematics control unit ~~that interfaces with the computer device on the vehicle, wherein the telematics control unit~~ that is configured to process ~~processes~~ the request based on [[in]] the telematics domain, wherein the one or more processors are part of the telematics control unit or are separate from the telematics control unit ~~context~~.

57. (**Currently Amended**) The mobile system of claim 56, wherein the one or more processors are configured to execute ~~further comprising~~ an event manager providing a multi-threaded environment for processing a plurality of requests in a plurality of different ~~contexts~~ domains across a plurality of overlapping or interleaved dialogs, wherein the event manager coordinates interaction between the parser and the telematics control unit to

enable the telematics control unit to process the request during the overlapping or interleaved dialogs.

58. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein the telematics control unit is configured to ~~processes the request in the telematics context in response to determining~~ determine whether ~~that~~ processing the request will ~~not~~ create a hazardous condition for the vehicle.

59. (**Currently Amended**) The mobile system of claim [[56]] 58, wherein ~~to process the request in the telematics control unit,~~ the telematics control unit is further configured to:

~~determines whether processing the request in the telematics context will create a hazardous condition for the vehicle;~~

provide ~~provides~~ interactive guidance to resolve the hazardous condition via an output device connected to the vehicle, wherein the telematics control unit provides the interactive guidance to resolve the hazardous condition in response to determining that processing the request ~~in the telematics context~~ will create the hazardous condition; and

process ~~processes~~ the request ~~in the telematics context~~ in response to the computer device ~~input device~~ receiving an input that manually overrides the hazardous condition determination.

60. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein ~~to process the request in the telematics control unit,~~ the telematics control unit is configured to:

~~receives~~ receive diagnostic information from one or more sensors connected to the vehicle, wherein the diagnostic information identifies one or more problems with the vehicle;

~~provides one or more interactive announcements or warnings~~ communicate the one or more problems via an output device connected to the vehicle ~~in response to receiving the diagnostic information that identifies the one or more problems with the vehicle~~; and

communicate with one or more systems that relate to resolving ~~mediates a solution to resolve~~ the one or more problems with the vehicle.

61. (**Currently Amended**) The mobile system of claim 60, wherein the one or more systems are related to ~~solution to resolve the one or more problems includes one or more of~~ scheduling service for the vehicle, summoning help to the vehicle over one or more communication links, providing remedial instructions that address the one or more problems, ordering one or more parts for the vehicle, or providing an estimated cost for the scheduling the service for the vehicle.

62. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein ~~to process the request in the telematics control unit,~~ the telematics control unit is configured to:

~~detects~~ detect an accident situation from information received from one or more sensors connected to the vehicle;

~~provides~~ provide one or more interactive announcements or warnings via an output device connected to the vehicle in response to detecting the accident situation, ~~wherein the telematics control unit provides the interactive announcements or warning to determine one or more of a~~ receive an input that is in response to the one or more interactive announcements or warnings, the input indicating a nature [[for]] of the accident situation or a condition [[for]] of a person in the vehicle; and

~~reports~~ communicate the nature of [[for]] the accident situation or the condition of [[for]] the person in the vehicle to emergency personnel over a communications link.

63. (**Currently Amended**) The mobile system of claim 62, wherein ~~to further process the request in the telematics control unit,~~ the telematics control unit is configured to establish ~~establishes~~ a voice channel over a communications link to enable communication between the person in the vehicle and the emergency personnel.

64. (**Currently Amended**) The mobile system of claim 56, wherein to process the request in the telematics control unit, the telematics control unit communicates over a communication link an indication that ~~summons help to the vehicle over one or more communication links in response to determining that~~ the one or more recognized words or

phrases include at least one predetermined panic word, panic phrase, emergency word, or emergency phrase ~~indicating that a crime is occurring~~.

65. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance for performing one or more maneuvers to operate the vehicle via an output device connected to the vehicle.

66. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein to process the request in the telematics control unit, the telematics control unit:

provides interactive routing information to coordinate a rendezvous with one or more other vehicles in communication with the vehicle; and

establishes a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles.

67. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein ~~to process the request in the telematics control unit,~~ the telematics control unit is further configured to:

~~detects~~ detect that the vehicle is approaching or has exceeded a travel limit for the vehicle;

~~provides~~ provide interactive guidance for returning to the travel limit via an output device connected to the vehicle in response to detecting that the vehicle is approaching or has exceeded the travel limit; and

determine whether ~~mediates a negotiation with a person in the vehicle~~ to extend the travel limit.

68. (**Currently Amended**) The mobile system of claim 86 [[56]], wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance relating to one or more tasks for a person in the vehicle based on a location of the vehicle.

69. (**Previously Presented**) The mobile system of claim 56, wherein the speech recognition engine recognizes the words or phrases in the natural language utterance using a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs.

70. (**Previously Presented**) The mobile system of claim 69, wherein the plurality of dictionary and phrase entries are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

71. (**Currently Amended**) A method for processing natural language utterances, comprising:

receiving a natural language utterance ~~at a speech unit connected to a computer device on a vehicle~~;

recognizing, at a speech recognition engine ~~connected to the computer device on the vehicle~~, one or more words or phrases in the natural language utterance;

identifying, at a parser, a telematics domain from the recognized one or more words or phrases of the natural language utterance;

~~interpreting, at a parser connected to the computer device on the vehicle, the recognized words or phrases, wherein interpreting the recognized words or phrases include identifying a telematics context for the natural language utterance~~;

formulating, at the parser, a request based on ~~in~~ the telematics domain ~~context from the interpreted words or phrases~~; and

communicating the request to ~~processing the request in the telematics context with~~ a telematics control unit that is configured to process the request based on the telematics domain ~~that interfaces with the computer device on the vehicle~~.

72. (**Currently Amended**) The method of claim 71, further comprising providing an event manager having a multi-threaded environment for processing a plurality of requests in a plurality of different domains ~~contexts~~ across a plurality of overlapping or interleaved

dialogs, wherein the event manager coordinates interaction between the parser and the telematics control unit to enable the telematics control unit to process the request during the overlapping or interleaved dialogs.

73. (**Currently Amended**) The method of claim 99 [[71]], wherein the telematics control unit is configured to determine whether ~~processes the request in the telematics context in response to determining that~~ processing the request will [[not]] create a hazardous condition for the vehicle.

74. (**Currently Amended**) The method of claim 99 [[71]], further comprising ~~wherein processing the request in the telematics control unit with the telematics control unit includes~~:

~~determining whether processing the request in the telematics context will create a hazardous condition for the vehicle;~~

providing interactive guidance to resolve the hazardous condition via an output device connected to the vehicle, wherein the telematics control unit provides the interactive guidance to resolve the hazardous condition in response to determining that processing the request in the telematics domain ~~context~~ will create the hazardous condition; and

processing the request ~~in the telematics context~~ in response to ~~the input device~~ receiving an input that manually overrides the hazardous condition determination.

75. (**Currently Amended**) The method of claim 99 [[71]], ~~wherein processing the request in the telematics control unit with the telematics control unit includes~~ further comprising:

receiving diagnostic information from one or more sensors connected to the vehicle, wherein the diagnostic information identifies one or more problems with the vehicle;

~~providing one or more interactive announcements or warnings~~ communicating the one or more problems via an output device connected to the vehicle ~~in response to~~

~~receiving the diagnostic information that identifies the one or more problems with the vehicle~~; and

~~mediating a solution to resolve~~ communicating with one or more systems that relate to resolving the one or more problems with the vehicle.

76. (**Currently Amended**) The method of claim 75, wherein the one or more systems are related to ~~solution to resolve the one or more problems includes one or more of~~ scheduling service for the vehicle, summoning help to the vehicle over one or more communication links, providing remedial instructions that address the one or more problems, ordering one or more parts for the vehicle, or providing an estimated cost for the scheduling the service for the vehicle.

77. (**Currently Amended**) The method of claim 99 [[71]], ~~wherein processing the request in the telematics control unit with the telematics control unit includes~~ further comprising:

detecting an accident situation from information received from one or more sensors connected to the vehicle;

providing one or more interactive announcements or warnings via an output device connected to the vehicle in response to detecting the accident situation [[,]] and receiving an input that is in response to the one or more interactive announcements or warnings, the input indicating ~~wherein the telematics control unit provides the interactive announcements or warning to determine one or more of~~ a nature [[for]] of the accident situation or a condition [[for]] of a person in the vehicle; and

~~reporting~~ communicating the nature [[for]] of the accident situation or the condition of [[for]] the person in the vehicle to emergency personnel over a communications link.

78. (**Currently Amended**) The method of claim 77, ~~wherein processing the request in the telematics control unit with the telematics control unit further includes~~ further comprising establishing a voice channel over a communications link to enable communication between the person in the vehicle and the emergency personnel.

79. (**Currently Amended**) The method of claim 71, further comprising communicating over a communications link an indication ~~wherein processing the request in the telematics control unit with the telematics control unit includes summoning help to the vehicle over one or more communication links in response to determining~~ that the one or more recognized words or phrases include at least one predetermined panic word, panic phrase, emergency word, or emergency phrase indicating that a crime is occurring.

80. (**Currently Amended**) The method of claim 99 [[71]], wherein processing the request in the telematics control unit with the telematics control unit includes providing interactive guidance for performing one or more maneuvers to operate the vehicle via an output device connected to the vehicle.

81. (**Currently Amended**) The method of claim 99 [[71]], wherein processing the request in the telematics control unit with the telematics control unit includes:

providing interactive routing information to coordinate a rendezvous with one or more other vehicles in communication with the vehicle; and

establishing a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles.

82. (**Currently Amended**) The method of claim 99 [[71]], ~~wherein processing the request in the telematics control unit with the telematics control unit includes~~ further comprising:

detecting that the vehicle is approaching or has exceeded a travel limit for the vehicle;

providing interactive guidance for returning to the travel limit via an output device connected to the vehicle in response to detecting that the vehicle is approaching or has exceeded the travel limit; and

determining whether ~~mediating a negotiation with a person in the vehicle~~ to extend the travel limit.

83. (**Currently Amended**) The method of claim 99 [[71]], wherein processing the request in the telematics control unit with the telematics control unit includes providing interactive guidance relating to one or more tasks for a person in the vehicle based on a location of the vehicle.

84. (**Previously Presented**) The method of claim 71, wherein the speech recognition engine recognizes the words or phrases in the natural language utterance using a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs.

85. (**Previously Presented**) The method of claim 84, wherein the plurality of dictionary and phrase entries are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

86. (**New**) The mobile system of claim 56, wherein the one or more processors are configured to receive the natural language utterance from a speech unit operatively coupled to a computer device in a vehicle, and wherein the telematics domain is a query and response domain or is a control domain.

87. (**New**) The mobile system of claim 86, wherein the query and response domain includes a driving directions domain, a travel services domain, an entertainment scheduling domain, or a location-based marketing domain, and wherein the control domain includes a vehicle lighting control domain, a windshield wiper control domain, or a vehicle entertainment domain.

88. (**New**) The mobile system of claim 86, wherein the parser is configured to identify the telematics domain further from a context of the natural language utterance.

89. (**New**) The mobile system of claim 88, wherein the parser is configured to determine the context based on a user profile associated with the natural language utterance or based on a previously received natural language utterance.

90. (**New**) The mobile system of claim 89, wherein the user profile includes information related to a user's vehicle type, vehicle location, special vehicle equipment, commonly visited destination, address, place of business, interest, or formatting or presentation preference.

91. (**New**) The mobile system of claim 88, wherein the one or more processors are configured to determine, based on the context, whether the received natural language utterance is part of a query request or part of a command request, and wherein the telematics control unit is configured to retrieve information responsive to the query request or cause an action consistent with the command request.

92. (**New**) The mobile system of claim 86, wherein the parser is configured to identify, from among a plurality of domain agents, a domain agent that is associated with the telematics domain and is configured to formulate the request based on the identified domain agent, and wherein the telematics control unit is configured to process the request by executing the identified domain agent.

93. (**New**) The mobile system of claim 86, wherein the one or more processors are configured to receive a second natural language utterance from the speech unit, and wherein the parser is configured to identify a non-telematics domain from one or more words or phrases of the second natural language utterance.

94. (**New**) The mobile system of claim 86, wherein the one or more processors are further configured to receive a non-voice input from the computer device in the vehicle, wherein the parser is configured to formulate the request further based on the non-voice input.

95. (**New**) The mobile system of claim 86, wherein the telematics control unit is operable to interface with a vehicle-related system, the vehicle-related system comprising a vehicle entertainment device, a communication device in the vehicle, a vehicle control system, a vehicle power management system, a vehicle sensor, a control device external to and in wireless communication with the vehicle, or a vehicle history storage device.

96. (**New**) The mobile system of claim 62, wherein the telematics control unit is configured to communicate medical information associated with the person in the vehicle over the communications link or via an output device in the vehicle.

97. (**New**) The mobile system of claim 86, wherein the system is configured to authenticate a user based on the received natural language utterance and grant vehicle access to the user in response to the authentication.

98. (**New**) The mobile system of claim 86, wherein the system is configured to receive an indication of fatigue of a user of the vehicle and to initiate a dialog to determine the user's fatigue level.

99. (**New**) The method of claim 71, wherein receiving the natural language utterance comprises receiving a natural language utterance at a speech unit operatively coupled to a computer device in a vehicle.

REMARKS

In response to the Non-Final Office Action mailed **December 7, 2011** (hereinafter "Office Action"), claims 56-68 and 71-83 have been amended, and claims 86-99 have been newly added. No claims have been cancelled. Therefore, claims 56-99 are pending. Support for the instant amendments is provided throughout the as-filed Specification. Thus, no new matter has been added. In view of the foregoing amendments and following comments, allowance of all the claims pending in the application is respectfully requested.

INFORMATION DISCLOSURE STATEMENT

An Information Disclosure Statement was filed by Applicants on December 6, 2010 along with a 12-page Form PTO-1449. *Only* page 1 of the 12-page Form PTO-1449 was returned with the Office Action (and signed by the Examiner). Accordingly, Applicants respectfully request that the Examiner provide a signed and initialed copy of the *entire* Form PTO-1449 for the 12/06/10 IDS with the next Office Action.

REJECTIONS UNDER 35 U.S.C. § 103

Independent claims 56 and 71 and dependent claims 58 and 73 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over U.S. Patent Application Publication No. 2002/0069071 by Knockeart *et al.* (hereinafter "Knockeart") in view of U.S. Patent No. 6,910,003 to Arnold *et al.* (hereinafter "Arnold"). Office Action, at page 2. The remaining dependent claims stand rejected as follows:

- Dependent claims 57 and 72 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart and Arnold, further in view of U.S. Patent Application Publication No. 2004/0205671 to Sukehiro *et al.* (hereinafter "Sukehiro"). Office Action, at page 5;

- Dependent claims 59 and 74 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart and Arnold, further in view of U.S. Patent Application Publication No. 2002/0022927 to Lemelson *et al.* (hereinafter "Lemelson"). Office Action, at page 7;

- Dependent claims 60-62, 65, 75-77, and 80 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart and Arnold, further in view of U.S. Patent No. 7,902,969 to Obradovich (hereinafter "Obradovich '969"). Office Action, at page 10;

- Dependent claims 63 and 78 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart, Arnold, and Obradovich '969, further in view of U.S. Patent No. 7,406,421 to Odinak *et al.* (hereinafter "Odinak"). Office Action, at page 16;

- Dependent claims 64 and 79 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart and Arnold, further in view of U.S. Patent No. 6,275,231 to Obradovich (hereinafter "Obradovich '231"). Office Action, at page 17;

- Dependent claims 66 and 81 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart, Arnold, and Obradovich '969, further in view of U.S. Patent No. 6,362,748 to Huang (hereinafter "Huang"). Office Action, at page 19;

- Dependent claims 68 and 83 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart and Arnold, further in view of U.S. Patent No. 6,754,485 to Obradovich *et al.* (hereinafter "Obradovich '485"). Office Action, at page 20;

- Dependent claims 69 and 84 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart and Arnold, further in view of U.S. Patent No. 6,877,001 to Wolf *et al.* (hereinafter "Wolf"). Office Action, at page 22; and

- Dependent claims 70 and 85 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over the alleged combination of Knockeart, Arnold, and Wolf, further in view of U.S. Patent No. 7,478,036 to Shen *et al.* (hereinafter "Shen"). Office Action, at page 24.

Applicants disagree with the propriety of the alleged rejections. However, *solely* in an effort to expedite prosecution, and in no way acquiescing in the propriety of the alleged rejections, independent claims 56 and 71 have each been amended to clarify aspects of the invention.

Independent claim 56, for example, recites a mobile system for processing natural language utterances, in which one or more processors are configured to execute:

> a parser that identifies a telematics domain from the recognized one or more words or phrases of the natural language utterance.

Independent claim 71 (as amended) recites similar features.

Knockeart relates to a "combined manual and spoken interface for a telematics system" that "provides a mixed graphical/manual and oral/spoken interface to a user." Knockeart, at ¶¶ 4, 38. The interface provides a "menu of a number of choices . . . includ[ing] 'navigation,' 'traffic,' 'news,' and 'emergency.'" Knockeart, at ¶ 44. A user can "select the 'navigation' choice . . . by speaking a predefined word to select the indicated choice, here the word 'select.'" Knockeart, at ¶ 44. That is, Knockeart relates to an interface that requires "limited to small numbers of words or phrases that can be more easily distinguished by a voice recognizer." Knockeart, Abstract. To select the navigation choice, the interface in Knockeart requires the user to use a "predefined word or phrase" such as "select." Knockeart, at ¶¶ 42-44. Thus, Knockeart fails to teach or suggest identifying a telematics domain from one or more words or phrases of a natural language utterance. Moreover, Arnold does appear to cure at least the foregoing deficiencies of Knockeart. Thus, Knockeart and Arnold, taken either alone or in combination, fail to teach or suggest each and every feature of independent claims 56 and 71.

For at least the foregoing reasons, withdrawal of the rejection of independent claims 56 and 71 is earnestly sought.

Dependent claims 57-70 and 72-99 each ultimately depend from and add features to one of independent claims 56 and 71. Accordingly, for at least the reasons discussed above with respect to independent claims 56 and 71, dependent claims 57-70 and 72-99 are patentable over the alleged combination of Knockeart and Arnold.

CONCLUSION

Having addressed each of the foregoing rejections, it is respectfully submitted that a full and complete response has been made to the outstanding Office Action and, as such, the application is in condition for allowance. Notice to that effect is respectfully requested.

If the Examiner believes, for any reason, that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at the number provided.

Respectfully submitted,

Date: **May 7, 2012**

PILLSBURY WINTHROP SHAW PITTMAN LLP

By: /Chen Li/
Chen Li
Registration No. 67,448

Customer No. 00909

P.O. Box 10500
McLean, Virginia 22102
Direct Dial: 703.770.7741
Main: 703.770.7900
Fax: 703.770.7901

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS :	Robert A. KENNEWICK *et al.*	CONFIRMATION No.:	8125
SERIAL NUMBER :	12/705,311	EXAMINER:	Fariba Sirjani
FILING DATE :	February 12, 2010	ART UNIT:	2626
FOR :	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE		

SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, Applicants hereby make of record in the above-identified application the documents listed on the attached Form PTO-1449.

Applicants reserve the right to establish the patentability of the claimed invention over any of the information provided herewith, and/or to prove that this information may not be prior art, and/or to prove that this information may not be enabling for the teachings purportedly offered.

This Supplemental Information Disclosure Statement is being filed more than three months after the filing date of a national application other than a continued prosecution application under 37 C.F.R. § 1.53(d), or more than three months after the date of entry of the

national stage as set forth in 37 C.F.R. § 1.491 in an international application, and after the mailing date of the first Office Action on the merits, but before the mailing date of any of a Final Action under 37 C.F.R. § 1.113, a Notice of Allowance under 37 C.F.R. § 1.311, or an action that otherwise closes prosecution in the application. 37 C.F.R. § 1.97(c).

Please charge the $180.00 fee [37 C.F.R. § 1.17(p)] to Deposit Account No. 03-3975. In addition, please charge any fee deficiencies, or credit any overpayment, to Deposit Account No. 03-3975 as needed to ensure consideration of the Supplemental Information Disclosure Statement.

It is respectfully requested that the Examiner initial and return a copy of the enclosed PTO-1449.

Respectfully submitted,

PILLSBURY WINTHROP SHAW PITTMAN LLP

By /Chen Li/
Chen Li
Registration No.: 67,448

P. O. Box 10500
McLean, Virginia 22102
Telephone: (703) 770-7900
Telefax: (703) 770-7901

Customer No.
00909

Dated: May 7, 2012

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	1A	5,027,406	06/25/1991	Roberts *et al.*	381	43	
	1B	5,208,748	05/04/1993	Flores *et al.*	364	419	
	1C	5,357,596	10/18/1994	Takebayashi *et al.*	395	2.84	
	1D	5,577,165	11/19/1996	Takebayashi et al.	395	2.84	
	1E	5,608,635	03/04/1997	Tamai	364	449.3	
	1F	5,740,256	04/14/1998	Castello Da Costa *et al.*	361	94.7	
	1G	5,748,974	05/05/1998	Johnson	395	759	
	1H	5,839,107	11/17/1998	Gupta *et al.*	704	270	
	1I	5,897,613	04/27/1999	Chan	704	210	
	1J	5,960,397	09/28/1999	Rahim	704	244	
	1K	5,960,399	09/28/1999	Barclay *et al.*	704	270	
	1L	5,960,447	09/28/1999	Holt *et al.*	707	500	
	1M	5,995,928	11/30/1999	Nguyen *et al.*	704	251	
	1N	6,018,708	01/25/2000	Dahan *et al.*	704	244	
	1O	6,035,267	03/07/2000	Watanabe *et al.*	704	1	
	1P	6,049,607	04/11/2000	Marash *et al.*	379	410	
	1Q	6,067,513	05/23/2000	Ishimitsu	704	233	
	1R	6,085,186	07/04/2000	Christianson *et al.*	707	3	
	1S	6,108,631	08/22/2000	Ruhl	704	270	
	1T	6,167,377	12/26/2000	Gillick *et al.*	704	240	
	1U	6,173,266	01/09/2001	Marx *et al.*	704	270	
	1V	6,188,982	02/13/2001	Chiang	704	256	
	1W	6,208,964	03/27/2001	Sabourin	704	244	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO
	1X	1 320 043 A2	06/18/2003	Europe				
	1Y	1 646 037	04/12/2006	Europe				

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

	1Z	"Statement in Accordance with the Notice from the European Patent Office dated 1 October 2007 Concerning Business Methods (OJ EPO 11/2007, 592-593), XP002456252

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	2A	6,233,561	05/15/2001	Junqua *et al.*	704	277	
	2B	6,266,636	07/24/2001	Kosaka *et al.*	704	244	
	2C	6,269,336	07/31/2001	Ladd *et al.*	704	270	
	2D	6,278,968	08/21/2001	Franz *et al.*	704	3	
	2E	6,366,882	04/02/2002	Bijl *et al.*	704	235	
	2F	6,377,913	04/23/2002	Coffman *et al.*	704	8	
	2G	6,397,181	05/28/2002	Li *et al.*	704	256	
	2H	6,405,170	06/11/2002	Phillips *et al.*	704	270	
	2I	6,429,813	08/06/2002	Feigen	342	357.13	
	2J	6,510,417	01/21/2003	Woods *et al.*	704	275	
	2K	6,522,746	02/18/2003	Marchok *et al.*	379	406.03	
	2L	6,549,629	04/15/2003	Finn *et al.*	381	92	
	2M	6,587,858	07/01/2003	Strazza	707	102	
	2N	6,594,367	07/15/2003	Marash *et al.*	381	92	
	2O	6,606,598	08/12/2003	Holthouse *et al.*	704	275	
	2P	6,622,119	09/16/2003	Ramaswamy *et al.*	704	9	
	2Q	6,633,846	10/14/2003	Bennett *et al.*	704	257	
	2R	6,681,206	01/20/2004	Gorin *et al.*	704	243	
	2S	6,701,294	03/02/2004	Ball *et al.*	704	257	
	2T	6,704,708	03/09/2004	Pickering	704	235	
	2U	6,751,612	06/15/2004	Schuetze *et al.*	707	4	
	2V	6,757,544	06/29/2004	Rangarajan *et al.*	455	456.1	
	2W	6,832,230	12/14/2004	Zilliacus *et al.*	707	203	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	3A	6,868,380	03/15/2005	Kroeker	704	240	
	3B	6,912,498	06/28/2005	Stevens *et al.*	704	235	
	3C	6,934,756	08/23/2005	Maes	709	227	
	3D	6,954,755	10/11/2005	Reisman	707	10	
	3E	6,959,276	10/25/2005	Droppo *et al.*	704	226	
	3F	6,975,993	12/13/2005	Keiller	704	275	
	3G	6,983,055	01/03/2006	Luo	381	313	
	3H	7,003,463	02/21/2006	Maes *et al.*	704	270.1	
	3I	7,024,364	04/04/2006	Guerra *et al.*	704	270	
	3J	7,035,415	04/25/2006	Belt *et al.*	381	92	
	3K	7,043,425	05/09/2006	Pao	704	211	
	3L	7,054,817	05/30/2006	Shao	704	270	
	3M	7,058,890	06/06/2006	George *et al.*	715	728	
	3N	7,069,220	06/27/2006	Coffman *et al.*	704	275	
	3O	7,072,834	07/04/2006	Zhou	704	244	
	3P	7,107,210	09/12/2006	Deng *et al.*	704	226	
	3Q	7,110,951	09/19/2006	Lemelson *et al.*	704	270	
	3R	7,143,037	11/28/2006	Chestnut	704	251	
	3S	7,165,028	01/16/2007	Gong	704	233	
	3T	7,184,957	02/27/2007	Brookes *et al.*	704	246	
	3U	7,197,069	03/27/2007	Agazzi *et al.*	375	233	
	3V	7,206,418	04/17/2007	Yang *et al.*	381	92	
	3W	7,207,011	04/17/2007	Mulvey *et al.*	715	812	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	4A	7,236,923	06/26/2007	Gupta	704	9	
	4B	7,277,854	10/02/2007	Bennett *et al.*	704	257	
	4C	7,305,381	12/04/2007	Poppink *et al.*	1	1	
	4D	7,328,155	02/05/2008	Endo *et al.*	704	251	
	4E	7,337,116	02/26/2008	Charlesworth *et al.*	704	254	
	4F	7,366,669	04/29/2008	Nishitani *et al.*	704	256	
	4G	7,415,414	08/19/2008	Azara *et al.*	704	270	
	4H	7,447,635	11/04/2008	Konopka *et al.*	704	275	
	4I	7,472,020	12/30/2008	Brulle-Drews	701	211	
	4J	7,472,060	12/30/2008	Gorin *et al.*	704	240	
	4K	7,493,259	02/17/2009	Jones *et al.*	704	257	
	4L	7,516,076	04/07/2009	Walker *et al.*	704	275	
	4M	7,536,297	05/19/2009	Byrd *et al.*	704	10	
	4N	7,536,374	05/19/2009	Au	706	55	
	4O	7,558,730	07/07/2009	Davis *et al.*	704	235	
	4P	7,574,362	08/11/2009	Walker *et al.*	704	275	
	4Q	7,577,244	08/18/2009	Taschereau	379	218.01	
	4R	7,606,708	10/20/2009	Hwang	704	257	
	4S	7,676,365	03/09/2010	Hwang *et al.*	704	240	
	4T	7,676,369	03/09/2010	Fujimoto *et al.*	704	270	
	4U	7,729,918	06/01/2010	Walker *et al.*	704	275	
	4V	7,788,084	08/31/2010	Brun *et al.*	704	7	
	4W	7,831,433	11/09/2010	Belvin *et al.*	704	275	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	5A	7,873,523	01/18/2011	Potter *et al.*	704	275	
	5B	7,917,367	03/29/2011	Di Cristo *et al.*	704	270.1	
	5C	7,949,529	05/24/2011	Weider *et al.*	704	270	
	5D	7,949,537	05/24/2011	Walker *et al.*	704	275	
	5E	7,983,917	07/19/2011	Kennewick *et al.*	704	257	
	5F	8,015,006	09/06/2011	Kennewick *et al.*	704	236	
	5G	8,060,367	11/15/2011	Keaveney	704	247	
	5H	8,069,046	11/29/2011	Kennewick *et al.*	704	257	
	5I	8,073,681	12/06/2011	Baldwin *et al.*	704	9	
	5J	8,086,463	12/27/2011	Ativanichayaphong *et al.*	704	275	
	5K	8,112,275	02/07/2012	Kennewick *et al.*	704	240	
	5L	8,140,327	03/20/2012	Kennewick *et al.*	704	226	
	5M	8,140,335	03/20/2012	Kennewick *et al.*	704	257	
	5N	8,145,489	03/27/2012	Freeman *et al.*	704	257	
	5O	8,150,694	04/03/2012	Kennewick *et al.*	704	257	
	5P	8,155,962	04/10/2012	Kennewick *et al.*	704	257	
	5Q	2001/0049601	12/06/2001	Kroeker *et al.*	704	254	
	5R	2002/0015500	02/07/2002	Belt *et al.*	381	66	
	5S	2002/0035501	03/21/2002	Handel *et al.*	705	10	
	5T	2003/0016835	01/23/2003	Elko *et al.*	381	92	
	5U	2003/0110037	06/12/2003	Walker *et al.*	704	257	
	5V	2003/0115062	06/19/2003	Walker *et al.*	704	258	
	5W	2003/0120493	06/26/2003	Gupta	704	270.1	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	6A	2003/0144846	07/31/2003	Denenberg *et al.*	704	277	
	6B	2003/0158731	08/21/2003	Falcon *et al.*	704	231	
	6C	2003/0206640	11/06/2003	Malvar *et al.*	381	93	
	6D	2003/0212550	11/13/2003	Ubale	704	215	
	6E	2004/0098245	05/20/2004	Walker *et al.*	704	1	
	6F	2004/0167771	08/26/2004	Duan *et al.*	704	10	
	6G	2005/0137877	06/23/2005	Oesterling *et al.*	704	275	
	6H	2005/0143994	06/30/2005	Mori *et al.*	704	235	
	6I	2007/0033020	02/08/2007	(Kelleher) Francois *et al.*	704	226	
	6J	2007/0073544	03/29/2007	Millett *et al.*	704	277	
	6K	2007/0118357	05/24/2007	Kasravi *et al.*	704	10	
	6L	2008/0065386	03/13/2008	Cross *et al.*	704	270	
	6M	2008/0140385	06/12/2008	Mahajan *et al.*	704	9	
	6N	2009/0216540	08/27/2009	Tessel *et al.*	704	275	
	6O	2009/0271194	10/29/2009	Davis *et al.*	704	235	
	6P	2011/0112827	05/12/2011	Kennewick *et al.*	704	9	
	6Q	2011/0112921	05/12/2011	Kennewick *et al.*	705	26.1	
	6R	2011/0131036	06/02/2011	Di Cristo *et al.*	704	9	
	6S	2011/0131045	06/02/2011	Cristo *et al.*	704	249	
	6T	2011/0231182	09/22/2011	Weider *et al.*	704	9	
	6U	2011/0231188	09/22/2011	Kennewick *et al.*	704	236	
	6V	2012/0022857	01/26/2012	Baldwin *et al.*	704	9	
	6W	2012/0101809	04/26/2012	Kennewick *et al.*	704	9	
	6Y	2012/0101810	04/26/2012	Kennewick *et al.*	704	9	
	6Z	2012/0109753	05/03/2012	Kennewick *et al.*	705	14.58	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 320 043 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
18.06.2003 Bulletin 2003/25

(51) Int Cl.[7]: **G06F 17/30**

(21) Application number: **02258126.8**

(22) Date of filing: **26.11.2002**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **13.12.2001 GB 0129788**

(71) Applicant: **Hewlett-Packard Company**
Palo Alto, CA 94304 (US)

(72) Inventor: **Belrose, Guillaume**
Bristol BS2 8HT (GB)

(74) Representative: **Squibbs, Robert Francis**
Hewlett-Packard Limited,
IP Section,
Building 3
Filton Road
Stoke Gifford, Bristol BS 34 8QZ (GB)

(54) Multi-modal picture

(57) A system for presenting a multi-modal picture includes picture presentation equipment (37) for displaying an image (38) of the picture and for enabling a user to interact with the picture by selecting specific picture features and asking a specific query relating to the feature. A voice browser system (3) controlled according to dialog scripts (46) associated with the picture, determines an appropriate response having regard to the spoken user query and the picture feature being pointed to. Each picture can have multiple narrators associated with it and the can choose which narrator is currently active. Picture authoring apparatus is also provided.

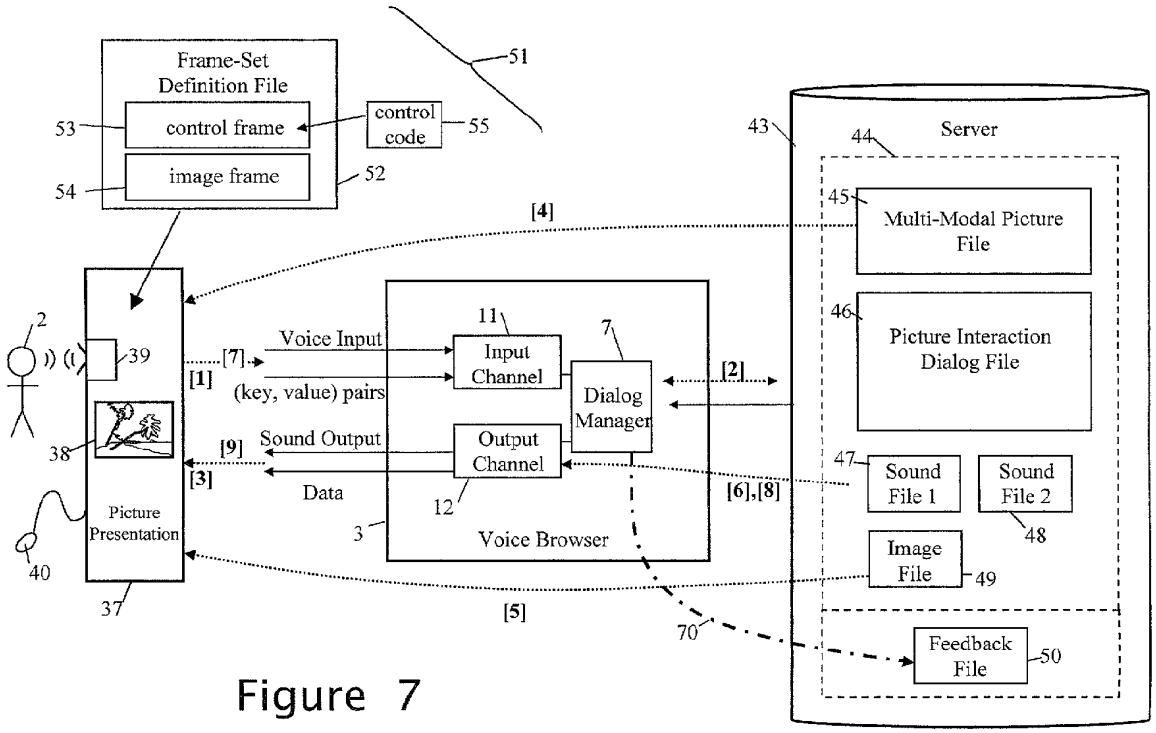


Figure 7

EP 1 320 043 A2

Google Exhibit 1019
Page 00878

Description

Field of the Invention

[0001] The present invention relates to multi-modal pictures with which a use can interact by spoken dialog exchanges.

Background of the Invention

[0002] In recent years there has been an explosion in the number of services available over the World Wide Web on the public internet (generally referred to as the "web"), the web being composed of a myriad of pages linked together by hyperlinks and delivered by servers on request using the HTTP protocol. Each page comprises content marked up with tags to enable the receiving application (typically a GUI browser) to render the page content in the manner intended by the page author; the markup language used for standard web pages is HTML (HyperText Markup Language).

[0003] However, today far more people have access to a telephone than have access to a computer with an Internet connection. Sales of cellphones are outstripping PC sales so that many people have already or soon will have a phone within reach where ever they go. As a result, there is increasing interest in being able to access web-based services from phones. 'Voice Browsers' offer the promise of allowing everyone to access web-based services from any phone, making it practical to access the Web any time and any where, whether at home, on the move, or at work.

[0004] Voice browsers allow people to access the Web using speech synthesis, pre-recorded audio, and speech recognition. Figure 1 of the accompanying drawings illustrates the general role played by a voice browser. As can be seen, a voice browser is interposed between a user 2 and a voice page server 4. This server 4 holds voice service pages (text pages) that are marked-up with tags of a voice-related markup language (or languages). When a page is requested by the user 2, it is interpreted at a top level (dialog level) by a dialog manager 7 of the voice browser 3 and output intended for the user is passed in text form to a Text-To-Speech (TTS) converter 6 which provides appropriate voice output to the user. User voice input is converted to text by speech recognition module 5 of the voice browser 3 and the dialog manager 7 determines what action is to be taken according to the received input and the directions in the original page. The voice input / output interface can be supplemented by keypads and small displays.

[0005] In general terms, therefore, a voice browser can be considered as a largely software device which interprets a voice markup language and generate a dialog with voice output, and possibly other output modalities, and / or voice input, and possibly other modalities (this definition derives from a working draft, dated September 2000, of the Voice browser Working Group of the World Wide Web Consortium).

[0006] Voice browsers may also be used together with graphical displays, keyboards, and pointing devices (e.g. a mouse) in order to produce a rich "multimodal voice browser". Voice interfaces and the keyboard, pointing device and display maybe used as alternate interfaces to the same service or could be seen as being used together to give a rich interface using all these modes combined.

[0007] Some examples of devices that allow multimodal interactions could be multimedia PC, or a communication appliance incorporating a display, keyboard, microphone and speaker/headset, an in car Voice Browser might have display and speech interfaces that could work together, or a Kiosk.

[0008] Some services may use all the modes together to provide an enhanced user experience, for example, a user could touch a street map displayed on a touch sensitive display and say "Tell me how I get here?". Some services might offer alternate interfaces allowing the user flexibility when doing different activities. For example while driving speech could be used to access services, but a passenger might used the keyboard.

[0009] Figure 2 of the accompanying drawings shows in greater detail the components of an example voice browser for handling voice pages 15 marked up with tags related to four different voice markup languages, namely:

- tags of a dialog markup language that serves to specify voice dialog behaviour;
- tags of a multimodal markup language that extends the dialog markup language to support other input modes (keyboard, mouse, etc.) and output modes (large and small screens);
- tags of a speech grammar markup language that serve to specify the grammar of user input; and
- tags of a speech synthesis markup language that serve to specify voice characteristics, types of sentences, word emphasis, etc.

[0010] When a page 15 is loaded into the voice browser, dialog manager 7 determines from the dialog tags and multimodal tags what actions are to be taken (the dialog manager being programmed to understand both the dialog and multimodal languages 19). These actions may include auxiliary functions 18 (available at any time during page processing) accessible through APIs and including such things as database lookups, user identity and validation, telephone call control etc. When speech output to the user is called for, the semantics of the output is passed, with any

Google Exhibit 1019
Page 00879

associated speech synthesis tags, to output channel 12 where a language generator 23 produces the final text to be rendered into speech by text-to-speech converter 6 and output to speaker 17. In the simplest case, the text to be rendered into speech is fully specified in the voice page 15 and the language generator 23 is not required for generating the final output text; however, in more complex cases, only semantic elements are passed, embedded in tags of a natural language semantics markup language (not depicted in Figure 2) that is understood by the language generator. The TTS converter 6 takes account of the speech synthesis tags when effecting text to speech conversion for which purpose it is cognisant of the speech synthesis markup language 25.

[0011] User voice input is received by microphone 16 and supplied to an input channel of the voice browser. Speech recogniser 5 generates text which is fed to a language understanding module 21 to produce semantics of the input for passing to the dialog manager 7. The speech recogniser 5 and language understanding module 21 work according to specific lexicon and grammar markup language 22 and, of course, take account of any grammar tags related to the current input that appear in page 15. The semantic output to the dialog manager 7 may simply be a permitted input word or may be more complex and include embedded tags of a natural language semantics markup language. The dialog manager 7 determines what action to take next (including, for example, fetching another page) based on the received user input and the dialog tags in the current page 15.

[0012] Any multimodal tags in the voice page 15 are used to control and interpret multimodal input/output. Such input/output is enabled by an appropriate recogniser 27 in the input channel 11 and an appropriate output constructor 28 in the output channel 12.

[0013] Whatever its precise form, the voice browser can be located at any point between the user and the voice page server. Figures 3 to 5 illustrate three possibilities in the case where the voice browser functionality is kept all together; many other possibilities exist when the functional components of the voice browser are separated and located in different logical/physical locations.

[0014] In Figure 3, the voice browser 3 is depicted as incorporated into an end-user system 8 (such as a PC or mobile entity) associated with user 2. In this case, the voice page server 4 is connected to the voice browser 3 by any suitable data-capable bearer service extending across one or more networks 9 that serve to provide connectivity between server 4 and end-user system 8. The data-capable bearer service is only required to carry text-based pages and therefore does not require a high bandwidth.

[0015] Figure 4 shows the voice browser 3 as co-located with the voice page server 4. In this case, voice input/output is passed across a voice network 9 between the end-user system 8 and the voice browser 3 at the voice page server site. The fact that the voice service is embodied as voice pages interpreted by a voice browser is not apparent to the user or network and the service could be implemented in other ways without the user or network being aware.

[0016] In Figure 5, the voice browser 3 is located in the network infrastructure between the end-user system 8 and the voice page server 4, voice input and output passing between the end-user system and voice browser over one network leg, and voice-page text data passing between the voice page server 4 and voice browser 3 over another network leg. This arrangement has certain advantages; in particular, by locating expensive resources (speech recognition, TTS converter) in the network, they can be used for many different users with user profiles being used to customise the voice-browser service provided to each user.

[0017] It is known to enhance pictures by providing associated speech annotations. It is an object of the present invention to provide further enhancements to such pictures.

Summary of the Invention

[0018] According to one aspect of the present invention, there is provided a system for presenting information concerning a picture to a user, the system comprising:

- a data store for holding responses, specific to said picture, in respect of specific user queries concerning particular picture features;
- a manually-operable feature-selection arrangement for enabling a user to select a feature in a displayed image of the picture, and for providing an output indication regarding what said particular feature, if any, the user has thereby selected;
- a voice dialog input-output subsystem including a speech recogniser for interpreting queries from a user;
- a control arrangement responsive to a user selecting a said particular feature and asking a specific query regarding that feature, to output the corresponding stored response.

[0019] Advantageously, each response is associated with a specified narrator and the user can opt to receive only the response of a selected narrator for either a current query or for all queries.

[0020] Preferably, provision is made for providing feedback to an author of the picture or its associated responses in the event that a user makes a query about a picture feature that is not a said particular picture feature or, although

Google Exhibit 1019
Page 00880

a query about such a feature, is not a said specific query for which a response is available.

[0021] According to another aspect of the present invention, there is provided a multi-modal picture specified by data held on at least one data carrier, this data comprising:

- picture image data for displaying a picture image;
- response data indicative of voice responses intended to be given to specific user queries concerning particular picture features of the picture;
- first control data for enabling a determination to be made as to which said particular feature in the picture image, if any, a user is selecting when using a selection arrangement to indicate a feature in the displayed image; and
- second control data for determining, on the basis of a spoken user query and on which said particular picture feature is selected by the user using the selection arrangement, which said response is to be used to reply to the user query.

[0022] According to a further aspect of the present invention, there is provided a multi-modal picture comprising a hard-copy picture image, and data held on at least one data carrier, this data comprising:

- response data indicative of voice responses intended to be given to specific user queries concerning particular picture features;
- first control data for enabling a determination to be made as to which said particular feature in the picture image, if any, a user is selecting when using a selection arrangement to indicate a feature of the image; and
- second control data for determining, on the basis of a spoken user query and on which said particular picture feature is selected by the user using the selection arrangement, which said response is to be used to reply to the user query.

[0023] The present invention also encompasses the overall process of generating a multi-modal picture and using it to convey information about particular picture features to a user. Thus, according to a still further aspect of the present invention, there is provided a method of conveying information about particular features in a picture, the method comprising the steps of:

(a) creating the following specifically-associated data:

- picture image data for displaying a picture image;
- response data indicative of voice responses intended to be given to specific user queries concerning particular picture features of the picture;
- first control data for enabling a determination to be made as to which said particular feature in the picture image, if any, a user is selecting when using a selection arrangement to indicate a feature in the displayed image; and
- second control data for determining, on the basis of a spoken user query and on which said particular picture feature is selected by the user, which said response is to be used to reply to the user query;

(b) using the image data to display an image of the picture;
(c) having a user use a manually-operated selection arrangement to select a feature in the displayed image and using the first control data to determine which said particular feature in the picture image, if any, the user is selecting;
(d) receiving and interpreting a spoken query from the user to determine if a said specific query is being asked; and
(e) using the second control data to determine, on the basis of the said particular feature determined as being selected in step (c) and the said specific query determined as being asked in step (d), which said response is to be used to reply and thereupon using the response data to output the corresponding voice response.

[0024] Similarly, according to a yet further aspect of the present invention, there is provided a method of conveying information about particular features in a hard-copy picture, the method comprising the steps of:

(a) creating the following specifically-associated data:

- response data indicative of voice responses intended to be given to specific user queries concerning particular picture features in said picture;
- first control data for enabling a determination to be made as to which said particular feature in the picture, if any, a user is selecting when using a selection arrangement to indicate a feature of the picture; and
- second control data for determining, on the basis of a spoken user query and on which said particular picture

Google Exhibit 1019
Page 00881

feature is selected by the user, which said response is to be used to reply to the user query;

(b) making the picture and data available to a user;
(c) having the user use a manually-operated selection arrangement to select a feature in the picture and using the first control data to determine which said particular feature in the picture, if any, the user is selecting;
(d) receiving and interpreting a spoken query from the user to determine if a said specific query is being asked; and
(e) using the second control data to determine, on the basis of the said particular feature determined as being selected in step (c) and the said specific query determined as being asked in step (d), which said response is to be used to reply and thereupon using the response data to output the corresponding voice response.

[0025] According to another aspect of the present invention, there is provided apparatus for authoring a multi-modal picture, comprising:

- a first tool for defining image hotspots associated with particular picture-image features;
- a second tool with speech recognition capability, for recording user responses input by voice, to user-specified queries each associated with a particular said picture-image feature; and
- means for automatically generating control data for determining, on the basis of a spoken user query and on which said particular picture feature is selected by a user, which said response is to be used to reply to the user query.

[0026] With respect to the above-mentioned feature regarding the provision of feedback on user queries for which responses are not available, a further independent aspect of the present invention provides a method of collecting user-interest information regarding a picture, the method comprising the steps of:

(a) presenting a picture image to a user, particular picture features in the image each having associated information that is presented to the user upon the user requesting such information;
(b) recognising a user information request, in respect of a picture feature, made by the user at least selecting that picture feature using a feature-selection arrangement; and
(c) in response to recognition of a user information request in step (b) concerning a picture feature that is other than a said particular picture feature, outputting an identifier of the concerned picture feature independently of the presented image.

[0027] This aspect of the invention may be implemented in the form of a system for collecting user-interest information regarding a picture, the system comprising:

- a data store for holding information concerning particular picture features;
- a feature-selection arrangement for enabling a user to select a feature in a displayed view of the picture, and for providing an output indication regarding what said particular feature, if any, the user has thereby selected;
- an information-output subsystem for outputting information about picture features to the user;
- a control arrangement for recognising a user information request, in respect of a picture feature, made by the user at least selecting that picture feature using the feature-selection arrangement, the control arrangement including:
  - first output means arranged to respond to a user making an information request concerning a said particular feature, by outputting the corresponding stored information via the information-output subsystem; and
  - second output means arranged to respond to a user making an information request concerning a picture feature that is other than a said particular picture feature, by outputting an identifier of the concerned picture feature for informing an author of the picture or its associated information regarding the unsatisfied information request.

Brief Description of the Drawings

[0028] A method and apparatus embodying the invention will now be described, by way of nonlimiting example, with reference to the accompanying diagrammatic drawings, in which:

. **Figure 1** is a diagram illustrating the role of a voice browser;
. **Figure 2** is a diagram showing the functional elements of a voice browser and their relationship to different types of voice markup tags;
. **Figure 3** is a diagram showing a voice service implemented with voice browser functionality located in an end-user system;

Google Exhibit 1019
Page 00882

- **Figure 4** is a diagram showing a voice service implemented with voice browser functionality co-located with a voice page server;
- **Figure 5** is a diagram showing a voice service implemented with voice browser functionality located in a network between the end-user system and voice page server; and
- **Figure 6** shows an example picture image of a multi-modal picture embodying the invention;
- **Figure 7** is a diagram of a system for presenting a multi-modal picture to a user;
- **Figure 8** is a diagram showing constituent dialog blocks of a picture interaction dialog file of the Figure 6 multi-modal picture;
- **Figure 9** is a diagram of apparatus for authoring a multi-modal picture;
- **Figure 10** shows, in respect of the Figure 6 image, a user seeking information from two picture features for which there is no associated information; and
- **Figure 11** shows the Figure 6 picture image enhanced with upper and lower information bars.

Best Mode of Carrying Out the Invention

[0029] In the following description, voice dialog interaction with a user is described based on a voice page server serving a dialog page with embedded voice markup tags to a multi-modal voice browser. Unless otherwise indicated, the foregoing description of voice browsers, and their possible locations and access methods is to be taken as applying also to the described embodiments ofthe invention. Furthermore, although voice-browser based forms of voice dialog services are preferred, the present invention in its widest conception, is not limited to these forms of voice dialog service system and other suitable systems will be apparent to persons skilled in the art.

Multi-Modal Picture

[0030] Figure 6 depicts a multi-modal picture comprising a displayed picture image 30 (here shown as being of a holiday island taken, for example, whilst the author of the picture was on holiday) with which a recipient (referred to below as the "user") can interact using multiple modalities and, in particular, by spoken dialogues and the use of a pointing arrangement, such as a cursor controlled by a mouse, stylus or keyboard keys, a touch-screen detection arrangement, etc. Thus, the user can query the picture using speech input, the query being either a general query about the picture or a specific query about a particular feature (item or area) of the picture indicated by use of the pointing arrangement (for example, the user uses a mouse to move a cursor over a picture feature and then clicks a mouse button). Of course, only certain picture features will have been set up to be queried and in the Figure 6 example, there are three such features, namely a coconut tree 31, a first person 32, and a second person 33; the dashed boxes around these features represent the "hotspot" image areas set up to encompass the features, the dashed lines generally not being visible (though displaying a hotspot boundary can be used as one way of indicating the hotspot to the user - a more typical way would be, in the case of the pointing arrangement being a mouse-controlled cursor, to change the cursor image as it moved into and out of a hotspot).

[0031] Processing functionality associated with the multi-modal picture is arranged to recognize one or more one or more general queries and, for each picture feature set up for selection, one or more specific queries. For each query set to be recognized, there is a corresponding response which is output to the user when the query is recognized, this output generally being in spoken form, generated using either speech synthesis or pre-recorded audio content. Thus, for example:

General Queries - available general queries can include picture location, the date of the picture was taken, and a description of the general subject of the picture. For instance:

User: "Describe the picture"
System: "This is a picture of the XYZ beach"
User: "What is the date?"
System: "The picture was taken November last year"
User: "What is the location?"
System: "XYZ beach is in Martinique, an island that is part of the French West Indies."

(In this and other dialog examples given in the present specification, the "System" is the functionality used to present the multi-modal picture to the user and to provide for multi-modal interaction).

Specific Queries - typically, the same specific queries are available for all the selectable picture features; example specific queries are: "What it is?", "Any story?", etc. For instance:

Google Exhibit 1019
Page 00883

User: "What is it"? "
System: "This is my cousin John".
User: "Any story?"
System: "John goes fishing quite often in the Caribbean Sea. One day, he just escaped a white shark attack."

[0032] The responses from the system can, of course, be more than just a simple single turn response and may involve a multi-turn structured dialog. Provision may also be made for different levels of content, for instance an initial level of public information and a lower level of private information only accessible under password control (or by some other security check procedure). Alternatively, there may be both a brief and a much fuller response .

[0033] A further possibility is to have responses from more than one narrator. Thus, a user can first obtain a response from one person associated with the picture and then a response from a different person also associated with the picture. For example, the user receives a picture from his/her parents. In the picture, the user's mother appears to be talking to someone, but the user's father looks bored. The user interaction might therefore proceed as follows:

User: "Mother, who is she?" [User uses pointing arrangement to indicate a person in the picture].
System: [In pre-recorded voice of user's mother] "This is my colleague from work. We had such a good time."
User: "Father, can you please describe the picture?"
System: [In pre-recorded voice of user's father] "We went to this rather boring evening with your mother's colleagues. They ended up talking about clothes all the time."

[0034] From the foregoing general description of a multi-modal picture, it will be appreciated that the term "multi-modal picture" covers more than just the picture image and includes also the associated dialog and the behaviour of the picture in response to user input.

Example Implementation

[0035] Figure 7 depicts an example implementation of a system for presenting a multi-modal picture.

[0036] The multi-modal picture is specified by a set 44 of related files held on a server 43, these files specifying both the image associated with the multi-modal picture and also the picture interaction dialogue associated with the picture. At least the dialogue and image files of set 44 are specifically associated with each other- that is, the association of these files is prespecified (in advance of user interaction) with the contents of the dialogue file being specifically tailored to the picture represented by the image file of the same set 44.

[0037] User 2 has local picture-presentation equipment 37; in the present case, this equipment is standard processing platform, such as a PC or portable computer, arranged to run a graphical web browser such as Microsoft's Internet Explorer. Other implementations of the picture presentation equipment 37 can alternatively be provided such as a mobile phone with a graphical display and suitable programs.

[0038] The picture-presentation equipment 37 displays the picture image of the multi-modal picture on display 38 and also provides for the input and output audio data via audio functionality 39, this audio input and output being transferred to and from a voice browser 3 (here shown as located separately from the picture picture-presentation equipment - for example, a network-based voice browser - but alternatively, integrated with the picture-presentation equipment 37). The picture-presentation equipment also has an associated pointing arrangement 40 depicted in Figure 7 as a mouse for controlling a cursor movable over the displayed picture image; other forms of pointing arrangements for selecting features of the picture image are, of course, also possible such as the use of touch pads, tracker balls or joysticks for moving an image cursor; touch-sensitive displays and other arrangements (such as a matrix of infra-red beams immediately overlying the display) for enabling a user to use a finger or stylus to point directly at a feature of the displayed image; etc..

[0039] The voice browser 3 comprises input channel 11, dialog manager 7 and output channel 12. The voice browser provides interactive dialog with the user, via the picture-presentation equipment 37, in respect of the currently-presented multi-modal picture, this dialog being in accordance with picture interaction dialog data retrieved from server 43.

[0040] The Figure 7 voice browser 3, as well as providing for voice input and output exchanges with the picture-presentation equipment 37, is also arranged to receive input data from the equipment in the form of key,value pairs indicating, for example, the selection of a particular picture feature by the user using the pointing arrangement 40 of the picture-presentation equipment. This data input is used by the dialog manager 7 in determining the course of the dialog with the user. The voice browser 3 can also provide data back to the picture-presentation equipment 37.

[0041] The picture-presentation equipment 37, voice browser 3 and server 3 inter-communicate via any suitable communications infrastructure or direct links.

[0042] Considering in more detail the files involved in the presentation of a multi-modal picture, these comprise, in addition to the set 44 of files that specify a multi-modal picture, a set of generic files 51. The multi-modal picture files

Google Exhibit 1019
Page 00884

44 comprise:

- the picture file 45, this file including a source reference for the picture image file 49 to be displayed and map data defining the image hotspots;
- a picture interaction dialog file 46 containing dialog scripts;
- one or more sound files 47, 48 (such as " .wav" files) containing audio data;
- the image file 49 (such as a ".jpg" or ".gif' file) containing the image data to be displayed.

[0043] The generic files 51 can be stored locally in the picture-presentation equipment 37 or retrieved from a remote location such as the server 43. The generic files comprise:

- a frame-set definition file 52 defining two frames 53, 54 into which page files can be independently loaded; one frame 53 is used to hold a file 55 containing control code (the contents of this frame not being visible), and the other frame 54 being used to hold the picture file 45 for the multi-modal picture to be presented.
- the control code file 55 to be loaded into frame 53, the control code being in the form of a number of scripts the main purpose of which is to provide key,value pairs to the voice browser according to events detected by the browser software run by the picture-presentation equipment 37 - in particular, clicking on an image hotspot as defined in file 45 is arranged to trigger a corresponding script in the control code file 55 whereby to cause a corresponding key,value pair to be passed to the voice browser 3 to inform it that a particular picture feature (corresponding to the activated hotspot) has been selected by the user.

[0044] To use the generic files to present a particular multi-modal picture, it is necessary to provide a reference to the multi-modal picture. This reference can, for example, be manually input by the user into an initial form displayed in the image frame 54 and then used by a script to request the loading of a file into frame 54. Alternatively, the multi-modal picture reference can be included as data in a query string attached to the URL of the frame-set definition file 52 (this URL and query string being, for example, provided to the user by the author of the multi-modal picture); in this case, in response to a request for the frame-set definition file 52, server-side code could, for example, extract the data from the query string and place it in the file source reference in the definition line for frame 54 in the frame-set definition file before that file is returned to the user.

[0045] In the present example, the multi-modal picture reference used to initiate presentation of the multi-modal picture is a reference to the picture interaction dialog file 46 to be loaded into the voice browser, rather than a reference to the picture file 45 that is to be loaded into the image frame 54. Thus, the multi-modal picture reference, however obtained, is passed in a key,value pair to the voice browser 3; voice browser 3 thereupon retrieves the picture interaction dialog file 46 to the dialog manager 7 of the voice browser. The file 46 includes a reference to the picture image file 45 to be loaded into the frame 54 and this reference is returned to the picture-presentation equipment 37 where it is used to retrieve the picture file 45. Where the picture-presentation equipment 37 comprises a standard web browser, one way of achieving the above is to have the dialog file reference sent (either as a source reference in the frame-set definition file or by a script in, for example, file 55) to the voice browser in a request for a file to load into frame 54, the voice browser sending back the picture file reference as a redirection.

[0046] It will be appreciated that the multi-modal picture reference passed into the generic files could have been that of the picture file 45, the latter then being retrieved into frame 54 and including an "onLoad" event script for passing to the voice browser a reference to the interaction dialog file.

[0047] The various steps involved in presenting a multi-modal picture according to the Figure 7 implementation are summarized below with reference to the reference numerals in square brackets in Figure 7:

[1] the generic files 51 have been loaded into the picture-presentation equipment 37 and a multi-modal picture reference has been provided; as a result, a key,value pair including a reference to the picture interaction dialog file 46 is passed to the voice browser 3;

[2] voice browser retrieves the dialog file 46 from the server 43 and dialog manger 7 uses the file to control further interaction with the user;

[3] the first action taken by the dialog manager 7 under the control of dialog file 46 is to pass to the equipment the reference for the picture file 45 along with a voice greeting to the user;

[4] the picture file 45 (including a reference to the image file 49) is retrieved from server 43 and loaded into the image frame 54;

[5] the image file 49 is retrieved from the server 43 and displayed on display 38;

[6] meanwhile, the dialog manager 7 causes a sound (in sound file 47) to be played to the user to indicate that the picture is ready to receive user input (this sound can simply be an appropriate background sound such as, for the Figure 6 picture image, the sound of the sea);

Google Exhibit 1019
Page 00885

[7] the user queries the picture by voice input (and possibly also by pointing to a particular area of the picture, this being indicated by a corresponding key,value pair sent to the voice browser along with the user voice input);
[8] dialog manager 7 acknowledges the receipt of the user query by causing an acknowledgement sound (in sound file 48) to be played back to the user;
[9] the dialog manager 7, having determined the appropriate response to the user query, outputs this response.

Steps [8] and [9] are repeated as many times as required by the user. In due course the user asks to exit and the dialog is terminated by the dialog manager.

[0048] Figure 8 illustrates the contents of the picture interaction dialog file 46. This file contains a number of dialog blocks 60 to 73 that contain dialog elements and/or control structures relating to dialog progression. Thus, dialog block 60 provides the initial greeting and causes the picture file reference to be passed to the equipment 37 (in step [3] above). Block 61 defines the query grammar and represents a waiting state for the dialog pending the receipt of a query from the user.

[0049] Block 62 carries out an analysis of a recognized query to determine whether it is an exit request (if so, an exit dialog block 63 is entered), a generic request, or a specific request; generic and specific requests are further analyzed to determine the nature of the query (that is, what "action" - type of information - is being requested). For a general query, the available actions are, in the present example, "date", "description", and "location; for a specific query, the action types are, in the present example, "what" and "story". Depending on the outcome of the action analysis, the dialog manager proceeds to one of blocks 64 - 66 (for a general query) or one of blocks 67 and 68 (for a specific query). The analysis carried out by dialog block 62 is on the basis of voice input only.

[0050] If the query was a general one, then block 64 is used to answer a date query, block 65 is used to respond to a description query, and block 66 is used to respond to a location query.

[0051] If the query was a specific "what" query, block 67 determines the identity of the picture feature (object) involved using the key,value pair provided to the voice browser; depending on the object identity, the dialog manager proceeds either to a "what" dialog block 70 for a coconut tree or to a "what" dialog block 71 for a French tourist. Similarly if the query was a specific "story" query, block 68 determines the identity of the picture feature (object) involved using the key,value pair provided to the voice browser; depending on the object identity, the dialog manager proceeds either to a "story" dialog block 72 for a coconut tree or to a "story" dialog block 73 for a French tourist.

[0052] After a response is provided by any one of the dialog blocks 64-66 or 70-73, the dialog manager returns to dialog block 61.

[0053] An example dialog interaction based on the Figure 6 picture and the Figure 8 dialog structure is given below; in this example, there is one narrator called Roger.

Example Interaction:

[0054]

System: "This picture was sent to you by Guillaume. You can talk to Roger to ask for information about it."
User: "Roger, Describe the picture"
System: "This is a picture of the XYZ beach."
User: "What is the location?"
System: "This picture was taken in Martinique, an island located in the French West Indies."
User: "What is it" [and clicks on feature 31]
System: "This is a coconut tree. This kind of trees is very common in the Caribbean."
User: "Any story?"
System: "When I was little, I used to get coconuts from my garden. There is a sweet liquid inside that is very tasty. We use coconut to make ice creams and all sorts of nice desserts."
User: "What is it?" [and clicks on 32]
System: "I guess this is a French tourist"
User: "Any story?"
System: "Tourism is one of the main source of revenue in Martinique. A lot of people come from France when it is winter over there."
User: "Exit"
System: "Thanks for using this digital picture frame."

[0055] The Appendix to the present description includes a detailed script example of the Figure 8 dialog interaction as well as the HTML source of a picture file 45.

Google Exhibit 1019
Page 00886

Multi-modal picture authoring tool.

[0056] Figure 9 illustrates apparatus for enabling an author 80 to author a multi-modal picture. The apparatus comprises a computing platform 81 providing a graphical user interface and running a tool, such as Microsoft's FrontPage product, for authoring mark-up language pages and for creating image hotspot maps. The apparatus further comprises a speech interface system 82 (here shown as a voice browser though other forms of speech interface system can be used). The speech interface system 82 permits the author 80 to interact with the apparatus by voice and is set up to recognize command words such as "Record".

[0057] Different people, known as narrators, can author different aspects of the same picture. The apparatus keeps a record of narrators known to it.

[0058] The apparatus is arranged to interact with one or more narrators to build up, in memory, the set of files 44 that specify a multi-modal picture, it being assumed that the picture image file 47 around which the multi-modal picture is to be built has already been loaded into memory (for example, from a digital camera) and is displayed via the graphical user interface (GUI) of platform 81. The process of building the required files is controlled by a top-level authoring program 90 that has three main steps 91-93 as will be more fully explained below.

Identifying the narrator - (step 91)

[0059] The first step of the authoring program is to identify the current narrator. The narrator speaks his/her name into the speech interface system 82; if the name is known to the apparatus, the system replies with a recognition greeting. However, if the narrator's name is not already known to the apparatus, the system asks the narrator to create a new profile (basically, input his/her name and any other required information), using appropriate data collection screens displayed on the graphical user interface of the computing platform 81.

[0060] Example: the apparatus knows the following names: "Lawrence" and "Marianne".

Apparatus: "What is your name?"
Narrator: "Steve".
Apparatus: "Sorry, I do not know this name, please write it down."
Narrator: [inputs "Steve" via a data collection screen of the GUI].

[0061] The authoring program uses the narrator's name to customize a greeting dialog block of a template picture interaction dialog file 46.

Adding general information about the picture - (step 92)

[0062] After identification, the narrator can input general information concerning the picture image such as the date, the location or the description, via a spoken dialogue. In the following example, the command words recognized by the speech interface system 82 are shown in bold whilst the nature of information being recorded (corresponding to the query "action" type of Figure 8) is indicated by underlining. The words indicating the nature of the information are either pre-designated to the system (effectively limiting the classification of information to be input) or else the system can be arranged to analyze narrator "Record" commands to determine the nature of the information to be recorded.

Example:

[0063]

Narrator: **"Record** description"
Apparatus: [Plays a beep].
Narrator: "This is a picture of me and John fishing in the Caribbean sea". [The apparatus records this input, either directly as sound data or as a text data after the input has been subject to speech recognition by the system 82]
Narrator: **"Write** date".
Apparatus: [displays date capture screen on GUI].
Narrator: [inputs date information via GUI].
Narrator: **"Record** story"
Apparatus: [Plays a beep.]
Narrator: "This day, John was attacked by a white shark." [The apparatus records this input]

Google Exhibit 1019
Page 00887

[0064] The authoring program uses the input from the narrator to create corresponding dialog blocks, similar to those described above with reference to Figure 8, in dialog file 46.

Adding specific information - (step 93)

[0065] The narrator can also input information concerning a specific feature of the picture image. To do this, the narrator indicates the picture feature of interest by using the GUI to draw a "hotspot" boundary around the feature. The apparatus responds by asking the narrator to input a label for the feature via an entry screen displayed on the GUI.. The authoring program uses the input from the narrator to create the multi-media picture file 45 holding an image hotspot map with appropriate links to the control code scripts.

[0066] The narrator can then enter further information using the speech interface system or the GUI. The narrator can record or write multiple descriptions or stories for a single area of the picture, for example, to give different level of details.

Narrator:	**"Record** description".
Apparatus:	[Plays beep.]
Narrator:	"It is a whale" [The apparatus records this input]
Narrator:	**"Record** story"
Apparatus:	[Plays beep.]
Narrator:	"We saw this whale on the way to Dominica." [The apparatus records this input]
Narrator:	**"Record next"** (indicates that further details for the story are to be recorded).
Narrator:	"We were crossing the Guadeloupe channel when we saw it." [The apparatus records this input]

[0067] Again, the authoring program uses the input from the narrator to create corresponding dialog blocks; thus, for the above example, where a "whale" hotspot has been designated by the user, the authoring program generates a set of dialogs blocks: 'whaleDescription', 'whaleStoryl', 'whaleStory2', etc.

[0068] After the first narrator has finished inputting information, other narrators can enter information in the same manner.

User Feedback

[0069] It will be appreciated that the authoring of a multi-modal picture, and in particular the adding of the dialog data, can be quite involved. As a result, it is quite likely that the author will not always include information that a recipient may be interested in.

[0070] It is therefore useful to be able to monitor a user's interaction with a multi modal picture to see if the user tries to access missing information. For example, a user receives the Figure 6 multi modal picture from a friend. Two objects 35,36 (a building and an island) in the picture image intrigues the user (see Figure 9) who therefore clicks on each picture feature concerned and asks for more information ("What is it?"). Unfortunately, there are no hotspots associated with either picture feature and therefore in each case the voice browser comes back with the response "Sorry, there is no information about this item."

[0071] However, the coordinates corresponding to each picture feature the user clicked are known to the browser used to display the picture image and the control script (for example, in file 55) can be used to pass the coordinates as key,value pairs to the voice browser 3. At the same time, a user verbal query is also passed to the voice browser. The voice browser first determines whether the query is a general one and if it is, the voice browser ignores the received coordinate values; however, if the voice browser determines that the query is a specific one, then it determines from the key,value pairs received that the user has indicated a picture feature for which there is no corresponding information available. In this case, the voice browser logs the received coordinate values and the associated "action" type (in the example given above, "what" or "story") in a feedback file 50 that forms part of the set 44 of related files associated with the multimodal picture (see arrow 70 in Figure 7). For example, upon the user clicking on the island feature 36 in Figure 9 and asking "what is this?", the data (action='what'; coordx =400;coordy=300) is logged to file 50. Such logging functionality is, for example, provided by a further dialog block of Figure 8. The logged coordinates provide, together with an indication of the picture concerned, a picture-feature identifier that identifies the picture feature about which information has been requested by the user.

[0072] Alternatively or additionally to logging the "desired-information" feedback data in file 50, the author of the multi-modal picture can be sent a message (for example, an e-mail message) explaining the query from the user such as "John wants a description of this object.". This message includes a picture feature identifier that identifies the picture feature concerned. The picture feature identifier can take the form of explicit data items indicative of the picture concerned and the coordinates of the feature in the picture or may more directly indicate the feature by including either

Google Exhibit 1019
Page 00888

image data representing the relevant portion of the picture image or image data showing the picture image with a marking indicating the position in the image of the feature concerned (both such forms of picture-feature indication can also be included the feedback file 50 additionally or alternatively to the feature coordinates). The picture-feature indication need not be sent to the file 50 (or included in a message to the author) at the time the system detects that the user has asked for information about a non-hotspot picture feature; instead, the indication and related query can be temporarily stored and output, together with other accumulated similar indications and queries, either at a specified time or event (such as termination of viewing of the picture by the user) or upon request from the picture author.

[0073] The author can then decide whether or not to add in further information by adding an additional hotspot and additional dialogs (for example, using the above-described authoring apparatus).

[0074] Of course, the same general feedback process can be used where although a selected picture feature is associated with an existing hotspot, there is no existing query "action" corresponding to a user's query. Furthermore, a similar feedback process can be used where user queries are input by means other than speech (such as, for example, via a keyboard or by a hand-writing or stroke recognition system) and, indeed, where there are no explicit user queries such as when the selecting of picture feature is taken as a request for information about that feature.

Variants

[0075] Many variants are, of course, possible to the arrangements described above. For example, Figure 11 illustrates a variant form of the multi-modal picture in which the picture image 30 is accompanied by upper and lower information bars 100 and 101 respectively. The upper information bar 100 indicates the narrators associated with the picture whilst the lower information bar 101 indicates what types of general and specific queries are available for use. These information bars assists the user in appreciating what queries can be put and to whom. To further assist the user, speaking (or clicking on) a narrator's name is preferably arranged to indicate what hotspots are associated with that narrator and where these hotspots are located in the image - thus, in Figure 11, narrator "Vivian" has been selected and hotspot 32 is indicated on the picture image by a dashed hotspot boundary line. Query types used by "Vivian" can also be indicated by highlighting these types (in the Figure 11 example, "Vivian" has used general query type 'description' and specific query types 'what' and 'story'). Preferably, once a narrator has been selected, that narrator remains selected until a different (or all) narrators is subsequently selected, whereby with a narrator selected, only the responses of that narrator will be used in responding to the users' queries. However, it is also possible to arrange for selection of a narrator to be effective only for a single query. In this latter case, a convenient way of providing for user selection of a narrator to answer a query about a particular feature is for a list of narrators of responses about that feature to be displayed whenever the user points to that feature.

[0076] Instead of identifying narrators by name as shown in Figure 11, other forms of identifier can be used such as an image of each narrator.

[0077] As regards the generic control code, this can be provided in the form of a Java applet or any other suitable form and is not limited to the use of client-side scripts of the form described above. Furthermore, rather than the frame-set definition file 52 and control code file 55 being generic in form, they can be made specific to each multi-modal picture.

[0078] As already indicated, picture invocation can be initiated in a variety of ways. As a further example, picture invocation can be arranged to be effected by the creator sending the user a reference to the picture interaction dialog file, the user sending this reference to the voice browser to enable the latter to retrieve the dialog file; the dialog file is arranged to cause the return to the user of the frame-set definition file (or a reference to it) with the latter already including the correct source reference to the picture file as well as to the control code file.

[0079] Persons skilled in the art will appreciate that there are very many ways of implementing multi-modal pictures and the supporting functionality ranging from the multiple file approach described above to having just a single data file containing all the necessary data and arranged to be executed by specifically adapted software. It will also be appreciated that whilst in the described embodiments the course of the system-user interaction is controlled by the control logic embedded in the picture interaction dialog file and interpreted by the dialog manager 7 of the voice browser, it is possible to provide this functionality separate from the voice browser and the response scripts (thus, in general terms, the dialog block 62, 67 and 68 would be used by a separate control arrangement for determining, on the basis of the user voice and pointing-arrangement inputs, which of multiple stored responses is the appropriate one to play back to the user). Furthermore, it is not necessary to explicitly identify a selected feature from the coordinates output by the pointing arrangement as a separate step to choosing an appropriate response to a particular user query; thus, a picture feature could simply be specified in the control logic of the dialog interaction file by its coordinate values (of range of values) whereby this control logic tests coordinate values output by the pointing arrangement against coordinate values of particular features in the course of selecting a response to a user query.

[0080] With respect to the selection of a picture feature of interest to a user, in the described embodiments this has been carried out by means of a user-operated pointing arrangement whereby data input about the picture feature of interest is generated through a manual operation. In addition to the described arrangements for manually effecting

Google Exhibit 1019
Page 00889

feature selection, it is possible to use an arrangement in which the coordinates of the feature of interest are manually input using a keypad (a coordinate system being displayed as part of the image presented to the user, or being presented around the boundary of the display area). Another possible feature-selection arrangement is one based on specifying, via keypad input, a particular area of the display where either the display is divided into labelled areas or there is a direct mapping between keypad keys and display areas. It is also possible to label each feature of interest in the image with a reference and have the user effect feature selection by keypad input of the appropriate reference.

[0081] Further as regards the determination of which picture feature is being selected by a user, it should be noted that determining the picture feature of interest from the image coordinates or image area identified by whatever selection arrangement is being used, can be done in ways other than that described above in which image coordinates generated by the feature-selection arrangement are mapped to picture features using predetermined mapping data as described above. For example, the image can have data encoded into it that labels particular picture features, the pointing arrangement being arranged to read this label data when over the corresponding picture feature. Technology for embedding auxiliary data in picture image data in a way that does not degrade the image for the average human viewer is already known in the art.

[0082] The picture image can be a hard-copy image carrying markings (such as infra-red ink markings) intended to be read by a sensor of a suitable pointing arrangement whereby to determine what particular picture feature is being pointed to; the markings can be a pattern of markings enabling the pointing arrangement to determine the position of its sensor on the image (in which case, an image map can be used to translate the coordinates into picture features) or the markings can be feature labels appropriately located on the image. Even without special markings added to the image, the image can still be a hard-copy image provided the image is located in a reference position relative to which the pointing arrangement can determine the position of a pointing element (so that an image map can be used to translate pointing-element position into picture features). Other manually-operated selection arrangements, such as those based on explicit coordinate input via a keypad, can also be used.

[0083] Where image map data is used to translate image coordinates to picture features, the image map data can be held separately from the picture file and only accessed when needed; this facilitates updating of the image map data. A reference (e.g. URL) to the image map data can be included in the picture file or, where the image is a hard-copy image, in markings carried by the image.

[0084] Preferably, the described embodiments are applied to pictures of scenes and places in the real world such as a tourist might take with a camera. However, it is also possible to apply the embodiments to topographic pictures that are primarily intended to convey map-type information.

Google Exhibit 1019
Page 00890

APPENDIX

This Appendix forms an integral part of the description.

Example of Multi-Modal Picture file.

The following HTML file was generated using Microsoft FrontPage and is the Multi-Modal Picture file 45 for the picture depicted in Figure 6; the HTML code pulls in image file "beachPicture.jpg" and defines three "hotspots" which when clicked result in a call to a function (method) of the control code file 55 with the passing of a parameter value indicating which hotspot has been clicked on.

File Name: MMPicture1.htm

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=windows-1252">
<meta name="GENERATOR" content="Microsoft FrontPage 4.0">
<meta name="ProgId" content="FrontPage.Editor.Document">
<title>Frame Page</title>
</head>
<body>
<p><map name="FPMap0">
<area href="http://localhost/mmpic1/dpfBlank.htm"
onclick="parent.parent.scriptFrame.pushParams('object=coconutTree')"
target="invisible" coords="128, 84, 253, 236" shape="rect">
<area href="http://localhost/mmpic1/dpfBlank.htm"
onclick="parent.parent.scriptFrame.pushParams('object=frenchTourist')"
target="invisible" shape="rect" coords="204, 266, 219, 282">
<area href="http://localhost/mmpic1/dpfBlank.htm"
onclick="parent.parent.scriptFrame.pushParams('object=frenchTourist')"
target="invisible" shape="rect" coords="309, 257, 324, 269"></map>
<img border="0" src="beachPicture.jpg" align="center" usemap="#FPMap0"
width="569" height="375"></p>
</body>
</html>
```


Google Exhibit 1019
Page 00891

Example of Picture Interaction Dialogue file

[0085] The following is an example implementation of the Figure 8 picture interaction dialog file 46. In this example, there is one narrator called Roger. The user can say things like:"Roger, what is the date?".

```
<talkml>
    <!--variables declaration-->
    <create name="object"          expr="'undefined'"/>
    <create name="action"          expr="'what'"/>
    <create name="type"        expr="'generic'"/>
    <!-- Specify here the url of the display webserver-->
    <display-server url="localhost"/>

    <dialog name="welcome">
        <say> Welcome to the digital picture frame. </say>
        <!-- Display the Multi Modal picture of beach. -->
        <display-show url="MMPicture1.htm"/>
        <say>
            This picture was sent to you by Guillaume. You can talk to Roger to
            ask for information about it.
        </say>
        <goto next="getUserRequest"/>
    </dialog>

    <dialog name="getUserRequest">
        <play src="sound1.wav"/>
        <display-listen wait_after="2000"/>
        <speech wait_after="2000">
            sentence =  narrator question | question | exit {action='exit'};
            question =  genericQuestion {type='generic'} | specificQuestion
                        {type='specific'};
```


Google Exhibit 1019
Page 00892

```
        narrator = roger {narrator='roger'};
        roger    = "roger";
        genericQuestion = date {action='date'}| location {action='location'}|
                          description {action='description'};
        date     = "what is the date";
        location = "what is the location";
        description = "describe the picture";
        specificQuestion= what {action='what'} | story {action='story'};
        what     = "what is it";
        story    = "any story";
        exit     = "exit";
    </speech>

    <recognised>
        <play src="sound2.wav"/>
        <goto next="filterRequest"/>
    </recognised>

    <unrecognised>
        <play src="sound2.wav"/>
        <say> I am sorry, I have some trouble understanding you.</say>
        <retry/>
    </unrecognised>

</dialog>

<dialog name="filterRequest">
    <if cond="action='exit'">
        <goto next="exit"/>
    </if>
```


Google Exhibit 1019
Page 00893

```
        <if cond="type='generic'">
            <if cond="action='description'"><goto next="description"/></if>
            <if cond="action='date'"><goto next="date"/></if>
            <if cond="action='location'"><goto next="location"/></if>
        </if>

        <if cond="type='specific'">
            <if cond="action='what'">    <goto next="what"/>  </if>
            <if cond="action='story'">    <goto next="story"/></if>
        </if>

</dialog>


<dialog name="description">
        <say>
            This is a picture of the beach.
        </say>
        <goto next="getUserRequest"/>
</dialog>


<dialog name="date">
        <say>
            Guillaume took this picture during his last trip home to
            Martinique last November
        </say>
        <goto next="getUserRequest"/>
</dialog>


<dialog name="location">
```


Google Exhibit 1019
Page 00894

```
        <say>
            This picture was taken in Martinique, an island located in the
            French West Indies.
        </say>
        <goto next="getUserRequest"/>
    </dialog>

    <dialog name="what">
        <if cond="object='coconutTree'"><goto next="coconutTreeWhat"/></if>
        <if cond="object='frenchTourist'"><goto next="frenchTouristWhat"/></if>
        <say>
            Apparently, you did not click on any parts of the picture. Try
            again.
        </say>
        <goto next="getUserRequest"/>
    </dialog>


    <dialog name="story">
        <if cond="object='coconutTree'"><goto next="coconutTreeStory"/></if>
        <if cond="object='frenchTourist'"><goto next="frenchTouristStory"/></if>
        <say>
            Apparently, you did not click on any parts of the picture. Try
            again.
        </say>
        <goto next="getUserRequest"/>
    </dialog>

    <dialog name="frenchTouristWhat">
        <say>
            I guess this is a French tourist.
        </say>
```


Google Exhibit 1019
Page 00895

```
        <goto next="getUserRequest"/>
    </dialog>


    <dialog name="coconutTreeWhat">
        <say>
            This is a coconut tree. This kind of trees is very common in the
            Caribbean.
        </say>
        <goto next="getUserRequest"/>
    </dialog>


    <dialog name="frenchTouristStory">
        <say>
            Tourism is one of the main source of revenue in Martinique. A lot
            of people come from France when it is winter over there.
        </say>
        <goto next="getUserRequest"/>
    </dialog>


    <dialog name="coconutTreeStory">
        <say>
            When I was little,I used to get coconuts from my garden. There is
            a sweet liquid inside that is very tasty. We use coconut to make
            ice creams and all sorts of nice desserts.
        </say>
        <goto next="getUserRequest"/>
    </dialog>
```


Google Exhibit 1019
Page 00896

```
<dialog name="exit">
        <say> Thanks for using this digital picture frame. </say>
    </dialog>

</talkml>
```

Claims

1. A system for presenting information concerning a picture to a user, the system comprising:

   - a data store for holding responses, specific to said picture, in respect of specific user queries concerning particular picture features;
   - a manually-operable feature-selection arrangement for enabling a user to select a feature in a displayed image of the picture, and for providing an output indication regarding what said particular feature, if any, the user has thereby selected;
   - a voice dialog input-output subsystem including a speech recogniser for interpreting queries from a user;
   - a control arrangement responsive to a user selecting a said particular feature and asking a specific query regarding that feature, to output the corresponding stored response.

2. A system according to claim 1, wherein image-map data is associated with the picture image for mapping image coordinates to said particular features, the selection arrangement being arranged to use the image-map data to determine what picture feature is selected by the user.

3. A system according to claim 1, wherein said image includes label data positioned in the region of a said particular picture feature to indicate the identity of that feature, the selection arrangement being arranged to read the label data to determine what picture feature is selected by the user.

4. A system according to any one of claims 1 to 3, further comprising a display subsystem for displaying said image provided to it in the form of digital image data.

5. A system according to any one of claims 1 to 3, wherein the picture image is a hard-copy image.

6. A system according to any one of the preceding claims, wherein the data store is arranged also to hold responses concerning general queries that are not associated with any particular picture feature, the control arrangement being arranged to respond to the user voice input of a general query by returning the appropriate response.

7. A system according to any one of the preceding claims, wherein the voice interface subsystem and control arrangement are arranged to cooperate in recognising multiple different queries in respect of a particular picture feature selected by the user.

8. A system according to claim 1, wherein each of at least some of the responses is associated with a specified narrator, the system being arranged to permit a user to receive only the response of a user-selected narrator in respect of at least one query.

9. A system according to claim 8, wherein user selection of a narrator is arranged to be effected by voice input, the control arrangement being arranged to respond to the selection of a particular specified narrator by user voice input, by using only the said responses associated with that narrator in providing a response to a user query.

10. A system according to claim 9, further comprising means for displaying, along with said picture image, identifiers of narrators associated with the picture.

11. A system according to claim 9, further comprising a display subsystem for displaying said image provided to it in

Google Exhibit 1019
Page 00897

the form of digital image data and identifiers of narrators associated with the picture; the display subsystem being arranged to respond to selection of a particular specified narrator by user voice input by indicating on the displayed image the said particular features for which responses are available concerning that narrator.

12. A system according to claim 8, further comprising a display subsystem for displaying said image provided to it in the form of digital image data, and identifiers of narrators associated with the picture; user selection of a narrator being arranged to be effected by the user using said selection arrangement to select a displayed narrator identifier, and the control arrangement being arranged to respond to the selection of a particular narrator by using only the said responses associated with that narrator in providing a response to a user query.

13. A system according to claim 12, wherein the display subsystem is arranged to respond to selection of a particular specified narrator to indicate on the displayed image the said particular features for which responses are available concerning the currently-selected narrator.

14. A system according to any one of the preceding claims, wherein said selection arrangement is a pointing arrangement usable by the user to point to a feature of interest in a displayed view of the picture.

15. A system according to any one of the preceding claims, further comprising a feedback arrangement arranged to respond to a user query concerning a picture feature that is other than a said particular picture feature, by outputting at least an identifier of the concerned picture feature for informing an author of the picture or its associated responses regarding the unsatisfied query.

16. A system according to any one of the preceding claims, further comprising a feedback arrangement arranged to respond to a user query that although concerning a said particular picture feature is not a said specific query, by outputting an indication of the query and an identifier of the concerned picture feature for informing an author of the picture or its associated responses regarding the unsatisfied query.

17. A system according to claim 15, wherein the feedback arrangement is further arranged to output an indication of the user query together with said picture feature identifier.

18. A system according to any one of claims 15 to 17, wherein the picture feature identifier arranged to be output by the feedback arrangement comprises an indication of the picture concerned and image coordinates of the picture feature.

19. A system according to any one of claims 15 to 17, further comprising means for converting digital picture data into said displayed image of the picture, the picture feature identifier arranged to be output by the feedback arrangement comprising digital data representing a corresponding portion of the image.

20. A system according to any one of claims 15 to 17, further comprising means for converting digital picture data into said displayed image of the picture, the picture feature identifier arranged to be output by the feedback arrangement comprising digital data representing the image with a marking indicating the position of the feature in the image.

21. A system according to any one of claims 15 to 20, wherein the feedback arrangement is operative to output said picture feature identifier to a log file.

22. A system according to any one of claims 15 to 20, wherein the feedback arrangement is operative to output said picture feature identifier in the form of an electronic message addressed to a person associated with the generation of the picture or its associated responses.

23. A multi-modal picture specified by data held on at least one data carrier, this data comprising:

- picture image data for displaying a picture image;
- response data indicative of voice responses intended to be given to specific user queries concerning particular picture features of the picture;
- first control data for enabling a determination to be made as to which said particular feature in the picture image, if any, a user is selecting when using a selection arrangement to indicate a feature in the displayed image; and
- second control data for determining, on the basis of a spoken user query and on which said particular picture

Google Exhibit 1019
Page 00898

feature is selected by the user using the selection arrangement, which said response is to be used to reply to the user query.

24. A multi-modal picture according to claim 23, wherein the first control data comprises image-map data mapping image coordinates to said particular features.

25. A multi-modal picture according to claim 23, wherein the first control data comprises label data arranged to be positioned in the displayed image in the region of each said particular picture feature to indicate the identity of that feature.

26. A multi-modal picture according to claim 23, wherein at least some of the said responses are associated with a narrator identified in the response data, the second control data enabling the determination of which said response is to be used to reply to the user query to be restricted to those responses associated with a said narrator that has been selected by the user.

27. A multi-modal picture according to any one of claims 23 to 26, wherein the picture is of a non-topographic real-world scene.

28. A multi-modal picture comprising a hard-copy picture image, and data held on at least one data carrier, this data comprising:

- response data indicative of voice responses intended to be given to specific user queries concerning particular picture features;
- first control data for enabling a determination to be made as to which said particular feature in the picture image, if any, a user is selecting when using a selection arrangement to indicate a feature of the image; and
- second control data for determining, on the basis of a spoken user query and on which said particular picture feature is selected by the user using the selection arrangement, which said response is to be used to reply to the user query.

29. A multi-modal picture according to claim 28, wherein the first control data comprises image-map data mapping image coordinates to said particular features.

30. A multi-modal picture according to claim 28, wherein the first control data comprises label data arranged to be positioned in or on the image in the region of each said particular picture feature to indicate the identity of that feature.

31. A multi-modal picture according to claim 28, wherein at least some of the said responses are associated with a narrator identified in the response data, the second control data enabling the determination of which said response is to be used to reply to the user query to be restricted to those responses associated with a said narrator that has been selected by the user.

32. A multi-modal picture according to any one of claims 28 to 31, wherein the picture is of a non-topographic real-world scene.

33. Apparatus for authoring a multi-modal picture, comprising:

- a first tool for defining image hotspots associated with particular picture-image features;
- a second tool with speech recognition capability, for recording user responses input by voice, to user-specified queries each associated with a particular said picture-image feature; and
- means for automatically generating control data for determining, on the basis of a spoken user query and on which said particular picture feature is selected by a user, which said response is to be used to reply to the user query.

Google Exhibit 1019
Page 00899

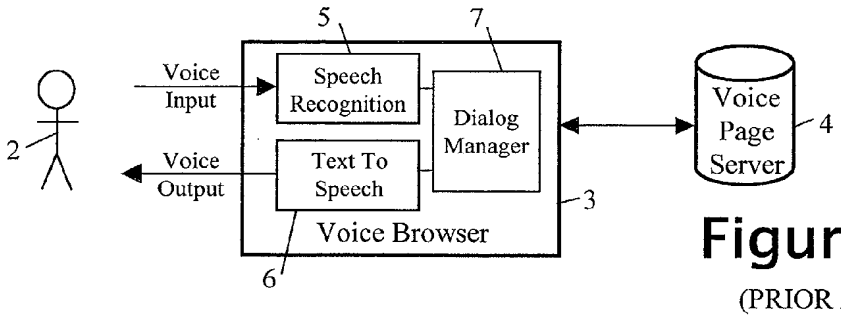


Figure 1

(PRIOR ART)

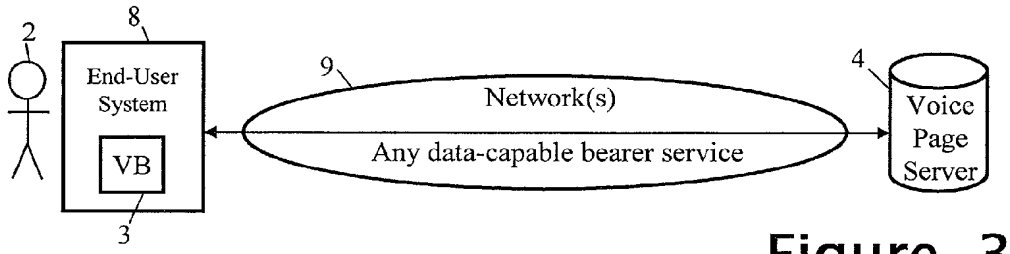


Figure 3

(PRIOR ART)

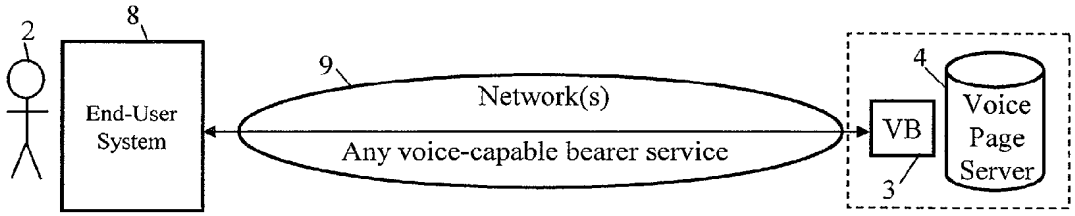


Figure 4

(PRIOR ART)

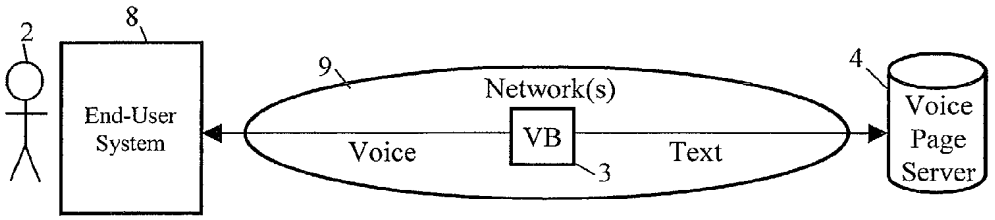


Figure 5

(PRIOR ART)

Google Exhibit 1019
Page 00900

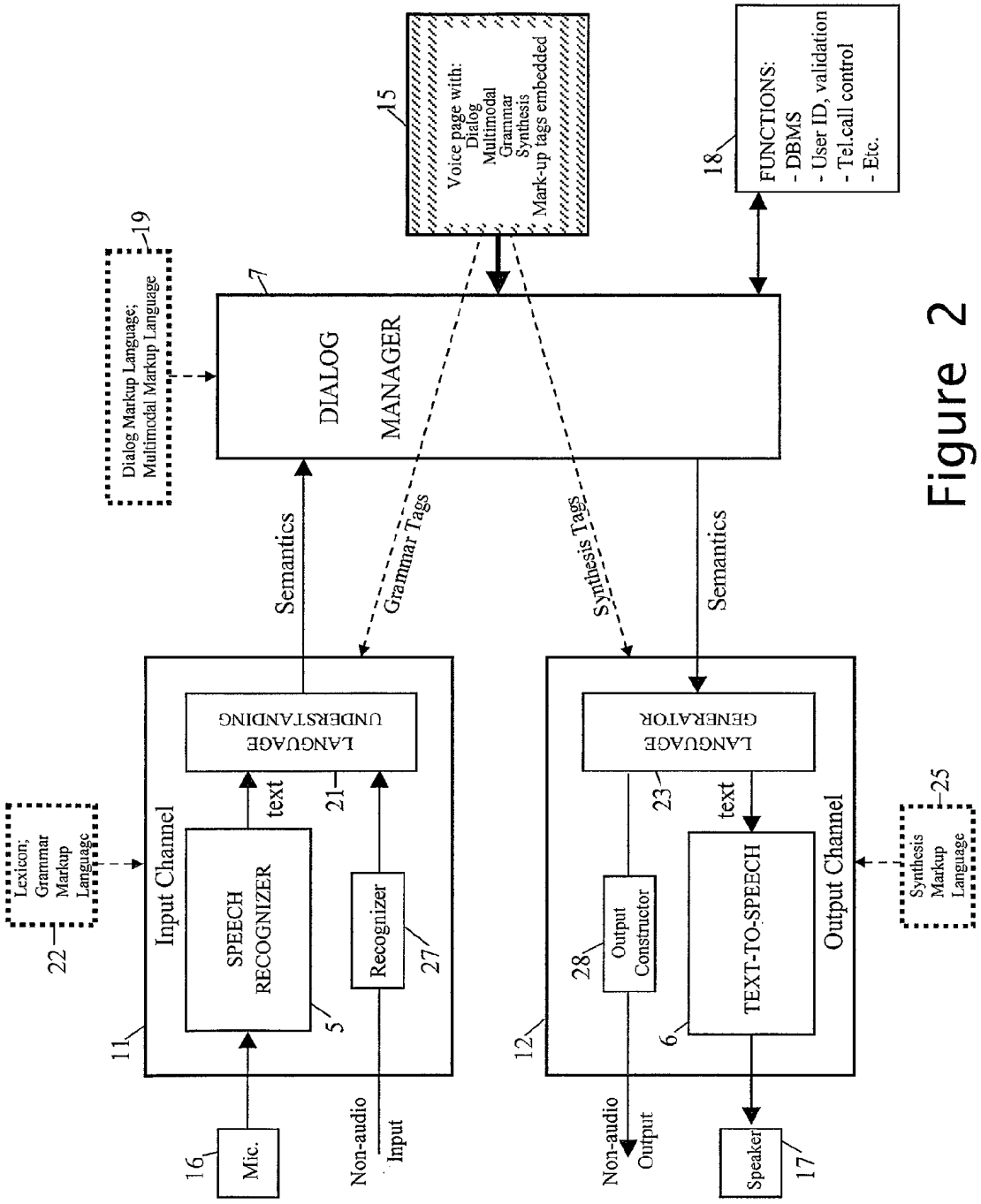


Figure 2
(PRIOR ART)

Google Exhibit 1019
Page 00901

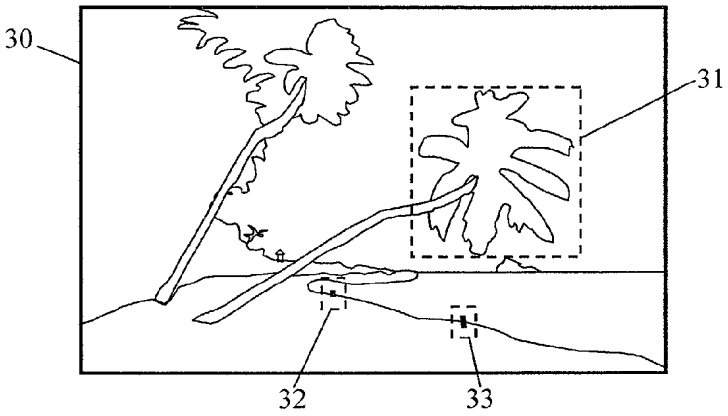


Figure 6

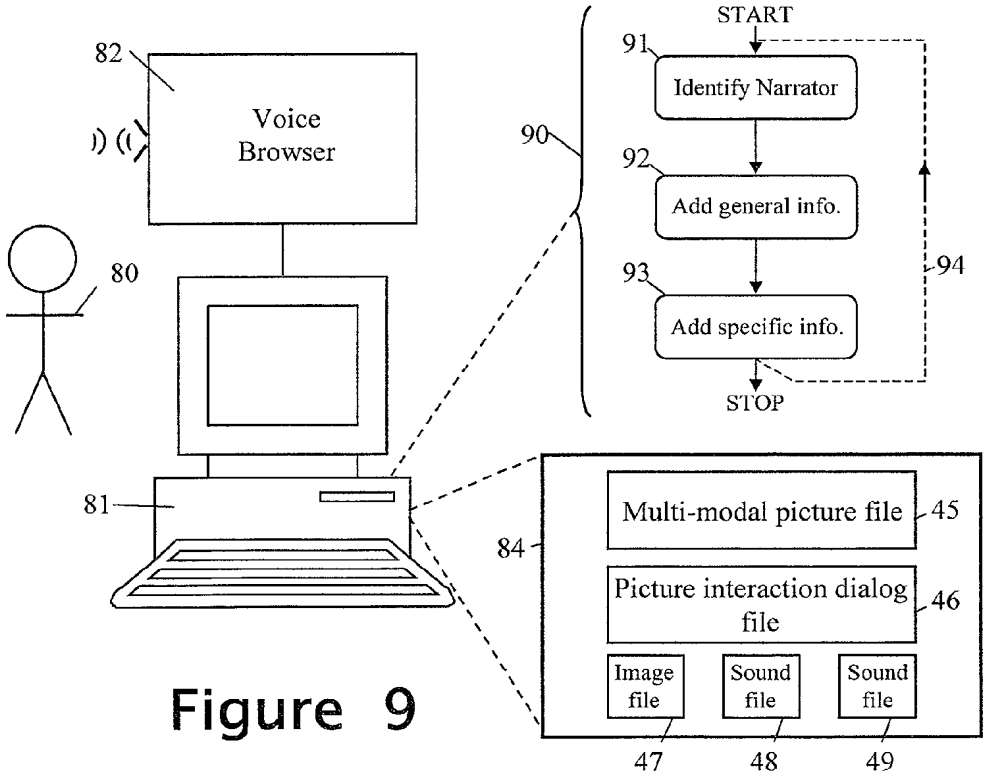


Figure 9

Google Exhibit 1019
Page 00902

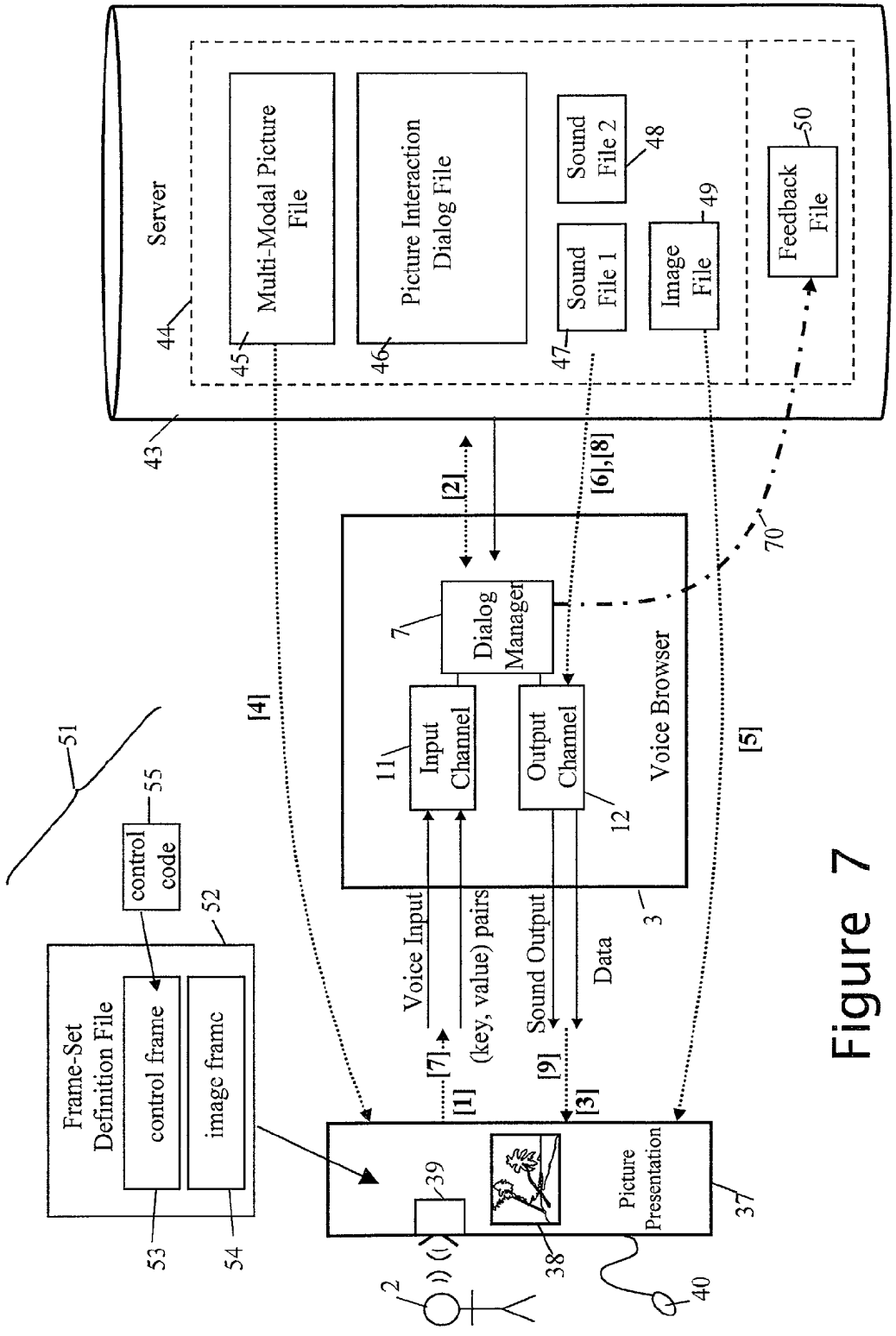
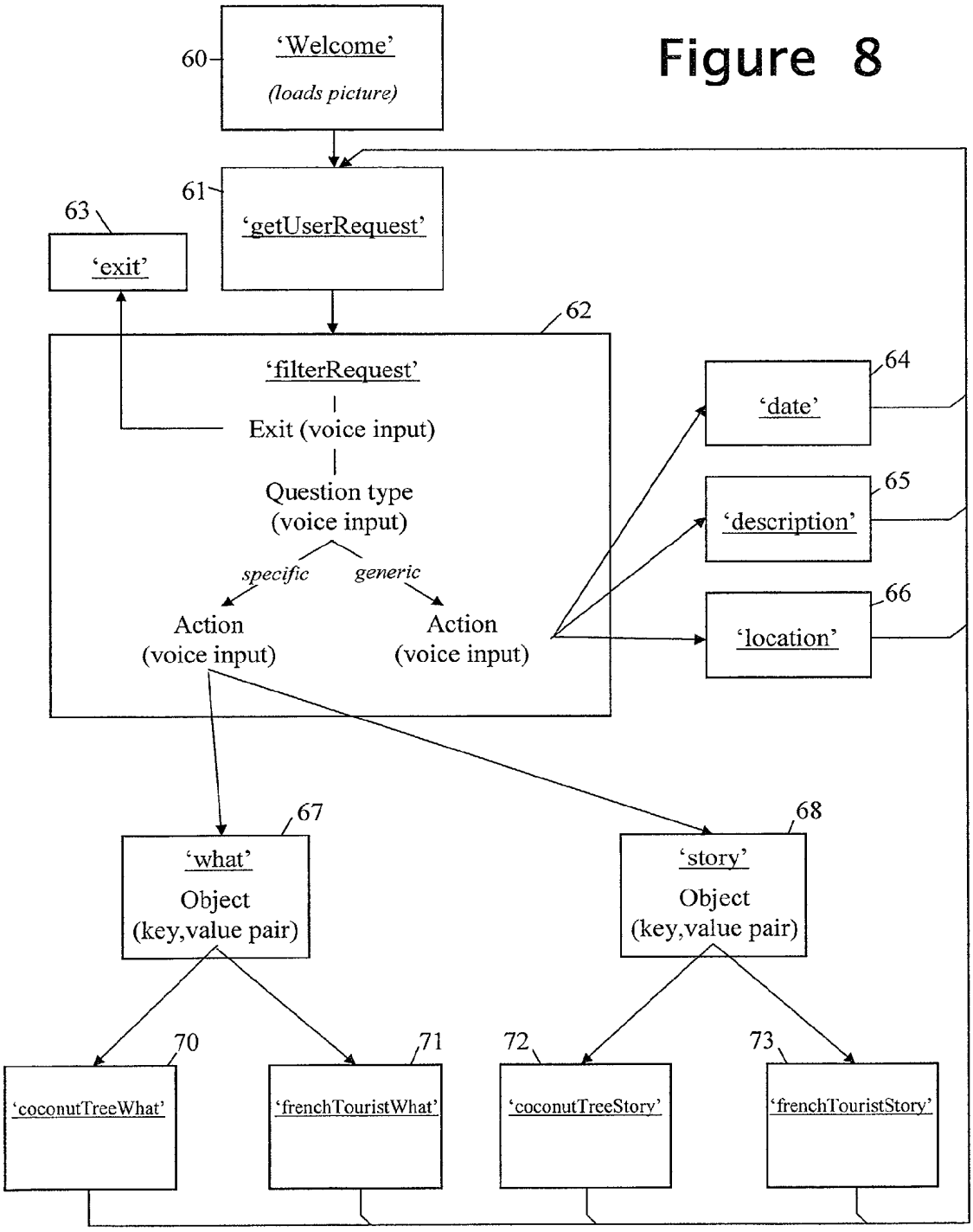


Figure 7

Google Exhibit 1019
Page 00903

Figure 8



Google Exhibit 1019
Page 00904

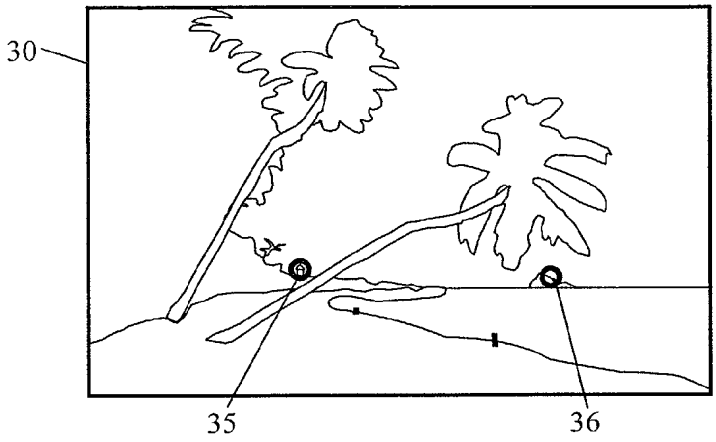


Figure 10

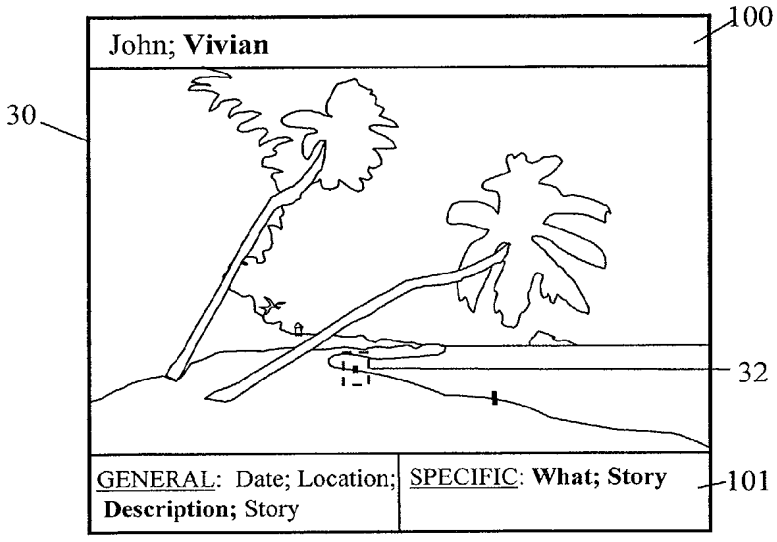


Figure 11

Google Exhibit 1019
Page 00905

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 646 037 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
26.03.2008 Bulletin 2008/13

(51) Int Cl.:
G10L 15/26 *(2006.01)*

(21) Application number: **05291989.1**

(22) Date of filing: **23.09.2005**

(54) Method and apparatus for enhancing speech recognition accuracy by using geographic data to filter a set of words

Verfahren und Vorrichtung für erhöhte Spracherkennungspräzision unter Verwendung geografischer Daten zur Filterung einer Wortgruppe

Procédé et appareil pour améliorer la précision de la reconnaissance vocale en utilisant des données géographiques pour filtrer un lot de mots

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR

(30) Priority: **29.09.2004 US 953671**

(43) Date of publication of application:
12.04.2006 Bulletin 2006/15

(73) Proprietor: **FRANCE TELECOM**
75015 Paris (FR)

(72) Inventors:
- **Waters, Keith**
  West Newton
  MA 02465 (US)
- **Simon, Benoît Benjamin**
  Boston
  MA 02114 (US)

(56) References cited:
EP-A- 0 661 688
EP-A- 1 233 407
US-B1- 6 230 132

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 1 646 037 B1

Google Exhibit 1019

Printed by Jouve, 75001 PARIS (FR)

Description

FIELD OF THE INVENTION

[0001] The invention relates generally to speech recognition systems and methods and, more particularly, to systems and methods for enhancing speech recognition accuracy by using geographic data to reduce the size of a set of words that must be checked to effect speech recognition.

BACKGROUND OF THE INVENTION

[0002] While visiting an unfamiliar city, a mobile handset user may wish to locate a business, landmark, street, or other place of interest. It would be quite advantageous if the user could place a voice call to a remote server at a designated telephone number, speaking his or her request for information into the handset. The handset would execute a speech recognition procedure, transforming the spoken request into a server query, and receiving the requested information from the remote server in the form of an electronic signal for producing audible speech. For example, a user in Manhattan may wish to locate the closest Neiman Marcus store. She dials a designated telephone number and speaks a request such as "where is the closest Neiman Marcus?" The handset uses speech recognition to convert the request into an electronic query for receipt by the remote server. Based upon the current location of the handset, the remote server accesses a database to retrieve the address and phone number of the nearest Neiman Marcus. The address and phone number are sent from the server to the user's handset in the form of an electronic signal, whereupon the handset converts the electronic signal into audible speech.

[0003] Unfortunately, there are significant hurdles to overcome if one wishes to use existing speech recognition algorithms to carry out the foregoing procedures. More specifically, it is a challenge to develop speech recognition algorithms, which provide a desired level of accuracy in certain operational environments, such as that of mobile telephonic devices. In general, the processing power and memory capabilities of mobile telephonic devices are limited by size and cost constraints, thereby placing limits on the extent to which the complexity of existing algorithms can be increased. Moreover, the existence of ambient noise severely degrades the performance of many speech recognition algorithms. In view of these shortcomings, prior art techniques have been developed to improve the performance of speech recognition algorithms in applications where processing power may be limited. These prior art techniques are directed toward enhancing a recognition model by utilizing improved acoustic models. The following definitions are presented herein for purposes of understanding and clarity:

[0004] *Lexical Model:* A lexical model breaks down individual words of a syntactic model (defined hereinafter) into a sequence of phonemes (also defined hereinafter).

[0005] *Acoustic Model:* An acoustic model breaks down the sequence of phonemes determined by the lexical model into a Markov model of cepstrum coefficients, also termed vector sequences. Cepstrum coefficients are generated by taking the Fourier Transform of the power spectrum of a speech signal.

[0006] *Recognition Model:* A recognition model is a set of rules and operations that are used by a recognition engine to recognize individual words. The recognition model includes a fully deployed speech recognition system which encompasses the acoustic model, the lexical model and the syntactic model as follows: recognition model = (acoustic model) x (lexical model) x (syntactic model).

[0007] U.S. Patent Application Publication No. US2003236099 discloses a technique for improving the accuracy of a speech recognition algorithm by selecting an appropriate acoustic model in accordance with the location of a mobile telephonic device ("mobile handset", or just "handset" for the sake of convenience). As indicated above, an acoustic model includes a set of cepstrum coefficient statistics for each phoneme in a language. Several different acoustic models can be precompiled, each representing a unique acoustical environment, such that a speech recognition system can select an acoustic model appropriate for a given set of environmental conditions. For example, each acoustic model is prepared by individually recording several hundred speakers in a controlled, model-specific laboratory environment that has known acoustical properties illustratively specified in terms of noise, echo, reverberation, and frequency response. Each model-specific laboratory environment is intended to represent a real-world environment, such as a warehouse, an office, a shopping mall, an airport, or the interior of an automobile. Hence, acoustic models incorporate properties of specific environments as well as characteristics of the human voice. The laboratory environment may include artificially generated sounds to simulate, for example, traffic, conversations, computer keyboard clicks, and forklift trucks.

[0008] Pursuant to this prior art technique, position information is retrieved from a location-based service (LBS), as defined hereinafter, based upon the current location of the mobile handset. The retrieved position information is then employed to select the acoustical properties of the environment surrounding the handset. An acoustic model that approximates the acoustical properties of the type of environment for the current location of the handset is selected from a set of previously prepared acoustic models. The selected acoustic model is then utilized to improve the accuracy of the speech recognition algorithm.

[0009] The technique disclosed in U.S. Patent Application Publication No. US2003236099 presents many practical shortcomings. Existing location-based services, using the Global Positioning System (GPS), will not

Google Exhibit 1019
Page 00907

determine a user's position to a requisite accuracy of, say, one to two meters, nor will such services determine a user's altitude with sufficient accuracy for many speech recognition applications. The impact of insufficient position accuracy is that, if a user is situated near the entrance of a building, it is not possible to accurately determine whether the user is, in fact, inside the building. The impact of insufficient altitude accuracy is that it is not possible to determine with certainty whether a user is within the building or standing on the roof. An acoustic model that is appropriate for an interior building lobby is unlikely to be totally appropriate for use in an outdoor setting. At the very least, selection of an inappropriate acoustic model based upon an inaccurate determination of a user's actual position will degrade the accuracy of many speech recognition algorithms.

[0010] Another shortcoming of the technique disclosed in U.S. Patent Application Publication No. US2003236099 is the need for creating an acoustic model. First of all, it is not feasible to create a unique acoustic model for every type of location. Secondly, even if an acoustic model is developed for a certain type of location, an actual location may not possess the acoustical characteristics embodied in the model. For example, due to variations in construction materials, furnishings, and room configurations, an acoustic model that attempts to represent a typical conference room, for example, may not adequately represent a specific conference room in a specific office building. Accordingly, what is needed is an improved technique for improving the accuracy of speech recognition algorithms.

[0011] EP1233407 recites a navigation system with speech recognition capacities for a moving vehicle. Whenever the distance between the vehicle and known active locations becomes greater then a given threshold, a pre-existing word list, corresponding to the active location, is acquired by the navigation system for subsequent use with the speech recognition engine.

[0012] US6230132 recites another navigation system with speech recognition functions. A lexicon is generated at periodic intervals with a limited number of words so as to facilitate the speech recognition functions. Furthermore, when a user needs to use the navigation system, he/she has to make a vocal request with the name of a location so that a lexicon can be generated.

[0013] EP0661688 recites a single device navigation apparatus, on board a car, with speed recognition functions. Pre-existing libraries are used and updated at regular intervals to facilitate the speech recognition engine.

SUMMARY OF THE INVENTION

[0014] The present invention is defined by the appended independent claims.

[0015] Other aspects and features of the present invention will become apparent from the following detailed description considered in conjunction with the accompanying drawings. It is to be understood, however, that the drawings are designed solely for purposes of illustration and not as a definition of the limits of the invention, for which reference should be made to the appended claims. It should be further understood that, unless otherwise indicated, the drawings are merely intended to conceptually illustrate the structures and procedures described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] In the drawings:

[0017] FIG. 1 is a diagrammatic representation of a geographic area of interest.

[0018] FIG. 2 is a data structure diagram setting forth a portion of an illustrative element database.

[0019] FIG. 3 is a data structure diagram setting forth an illustrative subset of words selected from the element database of FIG. 2.

[0020] FIG. 4 is a generalized information flow diagram showing a set of fundamental operations for the present invention.

[0021] FIG. 5 is a detailed block diagram setting forth information flow for a speech recognition system constructed in accordance with a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE DRAWINGS

[0022] The following terms used herein are defined as follows:

[0023] *Geographic Area of Interest:* Illustratively defined in terms of a zip code, group of zip codes, cell identification (Cell-ID) information, caller ID (Call-ID) information, the boundaries of a state, county, city, town, place, or other governmental or municipal subdivision, standard USGS (United States Geological Survey) grids, locations within a fixed range of latitude and longitude, or locations within a specified geometric boundary. The geographic area of interest may, but need not, be set automatically based upon the present location of a handset. The geographic area of interest may, but need not, be designated by a user based upon manual, verbal, or user-specified input. The geographic area of interest can be a coarse geographic area of interest or a refined geographic area of interest, with the latter being a portion of the former.

[0024] *E911:* As mandated by the FCC (Federal Communications Commission), a standard for a system by which wireless carriers are able to pinpoint the location of a mobile device in an emergency.

[0025] Element: Name of at least one of a street, a business, a merchant, an individual's name, a place, a point of interest, a transportation facility, an individual household, an activity, and a landmark.

[0026] *GPS*: Global Positioning System -- Enables geographic position determination at a land-based portable device equipped to receive signals from three or more orbiting GPS satellites.

[0027] *LBS:* Location-Based Service -- Any service

Google Exhibit 1019
Page 00908

that obtains information concerning the current geographical location of a mobile device user. Illustrative examples of location-based services include maps displayed in an automobile that track a motorists' position using GPS, or automatic localization services provided by dialing 911 on an E911-compliant mobile telephone network.

[0028] *Mobile Device:* Any portable device enabled for wireless communication, such as a wireless telephone, a personal digital assistant, or a laptop computer equipped with a wireless modem.

[0029] *Phoneme:* A phoneme represents a unique human uttered sound. Individual words, as pronounced by a human speaker, consist of a sequence of phonemes.

[0030] *Syntactic Model:* A set of rules and operations that specify allowed word sequences (phrases) of one or more words for subsequent use in a speech recognition procedure. Allowed word sequences are set forth in a list of static allowed phrases. In addition, the phrases can be optionally weighted by occurrence probabilities. Otherwise, word sequences can be in the form of stochastic or probabilistic lists, such as a Markov model of word sequences. Syntactic models are usually designed to be specific to a particular speech engine.

[0031] *Word:* A unit of language that native speakers can identify, and which communicates a meaning. Words are the building blocks from which phrases and sentences are constructed. Illustratively, for elements, "The Empire State Building" is the name of an element that contains four words. A street with the name "125th Street" includes the six words "one hundred and twenty fifth street". The street name "K Street" has the words "K" (which is a word for purposes of this invention) and "street".

[0032] The invention presumes that a very large element database is available for some sizeable, designated geographic area, such as a city, state, county, or country. One example of a known element database is TIGER (Topologically Integrated Geographic Encoding and Referencing system), available from the United States Census Bureau. TIGER compounds localized street names, rivers and railroads into a series of large regional databases.

[0033] The element database contains a set of words for elements at locations that are within the designated geographic area. In accordance with an aspect of the invention, enhanced speech recognition accuracy is provided by using geographic data to automatically select a subset of the words in the element database for use with a speech recognition procedure. The subset of words is selected by selecting only those words in the element database that are situated within a geographic area of interest. The geographic area of interest, either coarse or refined (depending on the circumstances, as explained below) is illustratively defined in terms of a zip code, group of zip codes, cell identification (Cell-ID) information, caller ID (Call-ID) information, the boundaries of a state, county, city, town, place, or other governmental or municipal subdivision, standard USGS (United States Geological Survey) grids, locations within a fixed range of latitude and longitude, or locations within a specified geometric boundary.

[0034] By selecting a subset of words, the total number of words used in the speech recognition procedure is substantially reduced. For example, a large element database containing street and business names used throughout all five boroughs of New York City can be reduced to a substantially shorter list of street and business names covering only the Grand Central Station area by specifying zip code 10017. Such a reduction in the total number of potential street names to be recognized, significantly improves speech recognition accuracy.

[0035] Pursuant to a further aspect of the invention, position information is obtained from an LBS based upon the current location of a mobile handset, illustratively using any of GPS coordinates, zip codes, and cellular base station triangulation. The obtained position information can be utilized to automatically define a coarse geographic area of interest. Let us say, as an example, that zip codes are used for this purpose. Thus, once the current location of the handset is determined, the zip code which includes that location automatically becomes the coarse geographic area of interest. Typically, the system can determine which localization method to use based on a knowledge of call configuration. For example, when a call is initiated, the Cell-ID can be used to determine the approximate location of the device. Alternatively, the application itself can choose which methods to employ, or the application can prompt the user to select which method to use. An illustrative prompt might include "*Do you want to use Cell-ID, or GPS to determine you location?*"

[0036] In accordance with a further aspect of the invention, an interface enables a user to refine the coarse geographic area of interest. For example, the user can select a specified geometrical shape within the coarse geographic area centered about the current location of the handset. This can be accomplished using key-based entries. Also, a graphical user interface can be used to draw a geometrical shape within the coarse geographic area. Additionally or alternatively, a Cell-ID can be specified within the geographic area zip code, thus defining a refined geographic area of interest for which the subset of words is to be selected. Combining automatic definition of a coarse geographic area of interest with additional, user-selectable geometric refinement is advantageous because it further reduces the total number of words that must be used in the speech recognition procedure. For example, a user-selectable rectangular region of 300 x 300 feet may contain only a few dozen elements that need to be recognized.

[0037] FIG. 1 is a diagrammatic representation of a geographic region of interest contained within a large area 100, such as a city. The current geographic location of a mobile device is determined to be point 401. In this illustrative example, the coarse geographic area of inter-

Google Exhibit 1019
Page 00909

est is defined by the zip code. Thus, the coarse geographic area of interest is defined by Zip code region 470 within which point 401 is located. A mobile device user has selected a specific radius 407 around point 401 for defining the refined geographic area of interest. Elements contained (in whole or in part) within this coarse geographic area include Street A 420, Street B 422, and a first building 406. Elements contained within this coarse geographic area of interest but which are not within this refined geographic area include Street C 424, second building 404, and third building 402.

[0038] An element database (FIG. 2, 500) includes a list of element names for at least one of streets, businesses, merchants, places, points of interest, transportation facilities, individual households, activities, or landmarks in large area 100 (FIG. 1). This database can be part of the LBS database. Each element name is associated with one or more corresponding parameters for identifying or associating the element with a refined geographic area of interest, and a parameter for identifying or associating the element with a coarse geographic area of interest. A set of phonemes corresponding to the element name can be computed in real time by recognition model compiler 213, using the outputted element name (501) in conjunction with a phonetizer. A phonetizer is a mechanism that provides phonemic transcriptions for each of a plurality of element names. Phonemic transcriptions comprise a phoneme, or a sequences of two or more phonemes, which indicate the manner in which an element name may be spoken. Illustratively, the element name "Madison" has a corresponding phonemic transcription "mäd è sun".

[0039] Element names are stored in an element name field 501. Each name in element name field 501 is associated with a corresponding parameter field (x,y) for the refined geographic area of interest 503, and a corresponding parameter field for the coarse geographic area of interest 505. These element names may or may not include elements that are within whatever refined geographic area of interest is selected by the user during operation of the invention (in this example, defined by radius 407 around point 401 of FIG. 1).

[0040] Refined geographic parameters are stored in the parameter field for the refined geographic area of interest 503 (FIG. 2). A refined geographic parameter sets forth a characteristic of an element in terms of how the refined geographic area of interest is defined. For example, if the refined geographic area is defined by the user as a shape (e.g., a circle) around the current location of the handset, the refined geographic parameters comprise (x,y) coordinates. In the illustrative example of FIG. 2, the refined geographic parameters are (x,y) coordinates determined with respect to an arbitrarily assigned reference point. When a first set and a second set of (x, y) coordinates are present in the parameter field for the refined geographic area of interest 503, this represents a line segment between the first set of (x,y) coordinates and the second set of (x,y) coordinates. Such a segment is used, for example, to define a street that extends from the first set of (x,y) coordinates to the second set of (x, y) coordinates.

[0041] Coarse geographic parameters are stored in the parameter field for the coarse geographic area of interest 505. A coarse geographic parameter sets forth a characteristic of an element in terms of how the coarse geographic area of interest is defined. For example, if the coarse geographic area of interest is defined in terms of one or more zip codes, the coarse geographic parameters in the parameter field for the coarse geographic area of interest 505 comprise one or more zip codes. For example, if Street A 420 (FIG. 1) crosses into two different zip code regions, then the parameter field for the coarse geographic area of interest 505 includes two zip codes, such as 10017 and 10036. However, other types of coarse geographic parameters could be employed to relate the element to a zip code system, such as standard USGS (United States Geological System) grids, regions within a fixed range of latitude and longitude, state boundaries, county boundaries, city boundaries, and/or boundaries of urbanized areas. For example, a lookup table could be used to determine the foregoing relationship.

[0042] Returning now to FIG. 1, and in accordance with one feature of the invention, only words corresponding to elements within the geographic area of interest are extracted from element database 500 (FIG. 2) to form word subset list 600 (FIG. 3). If the user defines the refined geographic area of interest as shown in FIG. 1, the limited set of words does not include Street C 424, second building 404, and third building 402. Use of this limited set of words enhances the accuracy of a subsequently performed speech recognition procedure because words pertaining to each and every element of large area 100 of FIG. 1 need not be considered.

[0043] FIG. 3 is a data structure diagram setting forth a portion of an illustrative word subset list 600. The names of one or more elements within the geographic area of interest are stored in an element name field 601. Optionally, word subset list 600 may include additional information, such as the probability that a particular element will be incorporated into a user's query based upon the distance of that element from the current position of the mobile device, so as to provide an enhanced level of performance.

[0044] FIG. 4 is a generalized information flow diagram describing a set of fundamental operations for the present invention. Geographic data acquisition mechanism 101 (the details of which are shown in FIG. 5) is equipped to collect geographic data corresponding to a mobile device's current geographic position and, based thereon, to determine and output a geographic area of interest. Geographic data includes, but is not limited to, any of: global positioning system (GPS) coordinates in latitude and longitude, and also altitude if needed, [x,y, z] coordinates, city name, place name, zip code area, geographic area, type of area (urban, suburban or coun-

Google Exhibit 1019
Page 00910

try), one or more cell identification (Cell-ID) parameters from a cellular network, or a caller identification (Call-ID) parameter from a terrestrial telephone network.

[0045] Word filtering mechanism 102 is equipped to receive a geographic area of interest from geographic data acquisition mechanism 101 and, based thereupon, select a subset of words from element database 500 (FIGs. 2 and 4). If a coarse geographic area of interest is being used, then the subset of words is selected by comparing coarse geographic parameters in the parameter field for the coarse geographic area of interest 505 corresponding to the elements in element name field 501 with coarse geographic parameters used to characterize the geographic area of interest. More specifically, if the coarse geographic area of interest is zip code 10017, then all the element names in element name field 501 will be selected which are associated with zip code 10017 in the parameter field for the coarse geographic area of interest 505. A similar operation is performed if a refined geographic area of interest is used. Of course, the parameter field for the refined geographic area of interest 503 will be used instead of the parameter field for the coarse geographic area of interest 505. The refined geographic parameters are combined with the coarse geographic parameters to enhance the word filtering mechanism 102 by using a two-step filtering process. A first step uses the parameter field for the coarse geographic area of interest 505, and a second step uses the parameter field for the refined geographic area of interest 503.

[0046] Word filtering mechanism 102 places the selected subset of words into a word subset list 600 (FIGs. 3 and 4). Speech recognition mechanism 103 (FIG. 4) is equipped to receive word subset list 600 (FIGS. 3 and 4) from word filtering mechanism 102 (FIG. 4) and, based thereupon, perform a speech recognition procedure. Illustratively, speech recognition mechanism 103 may be equipped to implement a speech recognition procedure by utilizing a (syntactic model) x (lexical model) x (acoustic model) paradigm in conjunction with Markov models. Alternatively or additionally, speech recognition mechanism 103 may utilize a neural network. More generally, the invention is applicable to any speech recognition mechanism that accepts and processes words or groups of words referred to as "grammar". Accordingly, word filtering mechanism 102 includes a grammar formatting mechanism that receives the word subset list 600 and formats the list into a grammar format usable by recognition model compiler 213.

[0047] Recognition model compiler 213 may or may not be required, depending upon the specific design characteristics of speech recognition mechanism 103. Recognition model compiler 213 parses (or analyzes) all of the entries in a formatted grammar, one after the other, and then, in one or more successive stages or "passes", constructs an output code readable by speech recognition mechanism 103. Since the speech recognition procedure is based upon a geographically refined subset of words, the overall accuracy of this procedure is enhanced relative to prior art designs. Speech recognition mechanism 103 can be implemented by a server or by a mobile handset.

[0048] A speech query from a mobile device user is received by an acoustic input mechanism 104 at the mobile device. Acoustic input mechanism 104 may be implemented using a microphone or other transducer equipped to transform acoustical vibrations into electronic signals. Acoustic input mechanism 104 feeds an electronic representation of user speech to speech recognition mechanism 103. Speech recognition mechanism 103 attempts to recognize one or more words in the electronic representation of user speech that are included in word subset list 600 retrieved from element database 500. If one or more words are recognized, speech recognition mechanism 103 sends an electronic message to an output mechanism 105 at the mobile device. The electronic message includes any recognized words in the speech query received from the mobile device user that were also included in word subset list 600. Output mechanism 105 converts the electronic message into audible and/or visual form for perception by the mobile device user. Illustratively, output mechanism 105 may be implemented using a loudspeaker coupled to audio amplifier circuitry and voice synthesizer circuitry for producing an audio message including one or more recognized words. Alternatively or additionally, output mechanism 105 may be implemented using a visual display device for displaying alphanumeric characters including one or more recognized words.

[0049] FIG. 5 is a detailed block diagram setting forth information flow for a speech recognition system constructed in accordance with an embodiment of the present invention. FIG. 5 shows details of geographic data acquisition mechanism 101. The depicted operations are performed either by a remote server 240 or by mobile device 250. Information flow commences upon receipt of a "start" signal at input mechanism 201 of mobile device 250. In practice, this "start" signal can be a request issued from a multimodal web page, an HTML page, or just a phone call. Alternatively, a user may place a call on a mobile handset to a designated telephone number to access a zip code form displayed on the mobile handset. The user fills in the zip code form using DTMF keypresses on the handset and validates the entered zip code by pressing DTMF digit "1" as a "start" signal. Basically, the "start" signal triggers a request for information.

[0050] Next, at mechanism 202, a decision is made as to whether the geographic area of interest will be determined locally (i.e., at mobile device 250), or at a remote location (i.e., at remote server 240). Depending upon the specifics of a given system application, this decision can be made by issuing a query to mobile device 250 in order to ascertain the configuration and capabilities of mobile device 250.

[0051] In cases where the geographic area of interest is determined locally, any one of two approaches may be employed. Pursuant to a first approach, mobile device

250 may include an optional GPS receiver 207, whereupon acquired GPS data are forwarded to a geographic-area-of-interest defining mechanism 204 (server side, see FIG. 5). Mechanism 204 uses the GPS data corresponding to the current location of mobile device 250 to obtain coarse geographic parameters for characterizing the coarse geographic area of interest. According to a second approach, zip code data, for example, can be entered by the user at input mechanism 201 and forwarded to geographic-area-of-interest defining mechanism 204. Illustratively, zip code data are entered into input mechanism 201 by sequentially pressing keys on a DTMF (dual-tone, multifrequency) keypad. Mechanism 204 uses the entered zip code data as a coarse geographic parameter for characterizing the geographic area of interest.

[0052] In cases where the geographic area of interest is determined remotely, mechanism 202 forwards the request received from input mechanism 201 to a location-based service 203 at server 240. Alternatively or additionally, GPS data from GPS receiver 207 are conveyed to location-based service 203, if needed. In response to the received request, location-based service 203 retrieves, for example, zip code information or cell identification (Cell ID) data for mobile device 250. Cell-IDs data can be retrieved in a mobile communications network using the Automatic Number Identification (ANI) retrieved from the mobile device 250. ANI is a service that provides the receiver of a telephone call with the number of the calling phone. The method of providing this information is determined by the service provider. Cell ID data specifies the geographic location of the wireless base station currently in communication with mobile device 250 by identifying at least one of the telephone number and the Internet Protocol (IP) address assigned to mobile device 250. For example, mobile device 250 may send to location-based service 203 Cell ID data in the form of an automated number identification (ANI) parameter. Optionally or alternatively, location-based service 203 determines the geographic location of mobile device 250 by performing a triangulation procedure among a plurality of wireless base stations. The Cell ID data, Caller ID data, and/or triangulation data retrieved by location-based service 203 (ultimately through the telecommunication provider) are forwarded to geographic-area-of-interest defining mechanism 204.

[0053] Optionally, in cases where the geographic area of interest is determined remotely, LBS service 203 utilizes at least one of Caller ID data and Cell ID data to access a zip code database 243. Based upon the Caller ID data or the Cell ID data, a zip code is retrieved from zip code database 243 and then sent to geographic-area-of-interest defining mechanism 204. Although the illustrative embodiment of FIG. 5 shows zip code database 243 and location-based service 203 as separate elements, those skilled in the relevant art will easily recognize that zip code database 243 could, but need not, be incorporated into location-based service 203. Likewise, although element database 500 and location-based service 203 are shown as separate elements, those skilled in the relevant art will easily recognize that element database 500 could, but need not, be incorporated into location-based service 203.

[0054] Geographic-area-of-interest defining mechanism 204 processes incoming geographic data, illustratively comprising at least one of incoming GPS data, zip code data, Caller ID data, or Cell ID data, to generate one or more geographic parameters defining the geographic area of interest for mobile device 250. For example, as discussed above, if the incoming data are based upon zip code data, then the output of mechanism 204 is a zip code. This parameter, e.g., a zip code, is used to specify a geographic area of interest. The defined geographic area of interest may combine a coarse geographic area of interest with a refined geographic area of interest, as is illustrated in the following example. One or more zipcode(s) may be deduced from GPS coordinates, thus defining a coarse geographic area of interest. The GPS coordinates themselves, associated with an area of uncertainty, define a refined geographic area of interest. Such a defined geographic area of interest may be used to implement the two-step filtering process described above.

[0055] Optionally, as indicated by the dotted lines of FIG. 5, a user at mobile device 250 may wish to further refine the geographic area of interest defined by geographic-area-of-interest defining mechanism 204. If so, the geographic area of interest defined by geographic-area-of-interest defining mechanism 204 is referred to as a coarse geographic area of interest, and the area of interest as refined by the user is referred to as a refined geographic area of interest. The refined geographic area of interest comprises a subset of the coarse geographic area of interest. Input mechanism 201 accepts user input setting forth a geometric specification of a refined geographic area of interest. User input is illustratively received over connection line 275 in the form of keypad entries, graphical objects, sketches, textual descriptions, or vocal input. Examples of geometric specifications include squares, rectangles, octagons, triangles, circles, and freehand sketches. Geometry specification refinement mechanism 205 accepts user input from input mechanism 201, and refines the set of geographic parameters generated by geographic-area-of-interest defining mechanism 204.

[0056] The one or more geographic parameters generated by geographic-area-of-interest defining mechanism 204, as optionally refined by geometry specification refinement mechanism 205, are received by word filtering mechanism 102. Word filtering mechanism 102 is enabled to carry out its functions when two conditions are met. Firstly, mobile device 250 should be executing a speech-enabled application where the user has issued a request for information. This requirement is met upon generation of a request from input mechanism 201. Secondly, the geographic area of interest should be defined.

Google Exhibit 1019
Page 00912

This requirement is met once geographic-area-of-interest defining mechanism 204 generates the above-described geographic parameters.

[0057] When enabled, word filtering mechanism 102 uses the one or more geographic parameters received from mechanism 204 or 205 to determine whether or not each of a plurality of elements in element database 500 is situated within the geographic area of interest. The element names stored in element name field 501 (FIG. 2) may include elements that are within the geographic area of interest, as well as elements that do not lie within the geographic area of interest. Accordingly, word filtering mechanism 102 (FIG. 5) compares the one or more received parameters defining the geographic area of interest against one or more parameters stored in at least one of parameter field for the refined geographic area of interest 503 or parameter field for the coarse geographic area of interest 505 (FIG. 2), to determine whether or not each of a plurality of elements in element database 500 (FIGs. 2 and 5) is situated within the geographic area of interest. For example, Street A is characterized in the parameter field for the refined geographic area of interest 503 by a first and a second set of (x,y) coordinates. Accordingly, Street A is conceptualized as including a straight line segment connecting the first and second sets of (x,y) coordinates. If any portion of this straight line segment intersects the geographic area of interest, then Street A is included in word subset list 600 (FIGs. 3 and 5).

[0058] Word filtering mechanism 102 appends to word subset list 600 only those element names within the geographic area of interest. Optionally, if the user has made any refinements to the coarse geographic area of interest, only elements that are within the refined geographic area of interest are appended to word subset list 600.

[0059] The word subset list 600 of FIGs. 3 and 5 represents an illustrative output format for word filtering mechanism 102. Alternatively or additionally, the output of word filtering mechanism 102 may be conceptualized as a computer-readable stream of words representing one or more elements in a geographic area of interest. This stream, illustratively in text format, is compatible with at least one of a speech recognition engine or a voice browser embedding speech recognition mechanism 103 (FIG. 5). Optionally, word filtering mechanism 102 may reformat at least one of word subset list 600 or the computer-readable stream of words into a format such as an XML document, to provide a subset of words compatible with speech recognition mechanism 103.

[0060] Speech recognition mechanism 103 may be implemented at server 240 or at mobile device 250. In cases where speech recognition is performed at mobile device 250, the output from word filtering mechanism 102 (illustratively in the form of a stream of words corresponding to grammar subset list 600, FIG. 3) is sent from server 240 to mobile device 250. Word subset list 600 is then compiled by the recognition model compiler 213 at mobile device 250 to provide a speech recognition model for use by a speech recognition mechanism 103 responsive to the acoustic input mechanism 104. Speech recognition mechanism 103 then forwards recognized speech to the output mechanism 105 for generating an electronic output signal representing the recognized speech. Such an arrangement is advantageous because it reduces the burden on wireless communication resources. Transmissions between the server and the mobile device are reduced because only word subset list 600 needs to be communicated to the mobile device. The speech recognition steps can be performed at the mobile device with reference to word subset list 600, thereby enhancing the speed and efficiency of the speech recognition process.

[0061] If speech recognition is performed at server 240, then recognition model compiler 213 and speech recognition mechanism 103 are implemented at server 240 instead of at mobile device 250, wherein the output of speech recognition mechanism 103 is sent from server 240 to output mechanism 105 at mobile device 250.

[0062] The operational principles of the configurations shown in FIGs. 4 and 5 are best illustrated with reference to several illustrative scenarios. These scenarios clarify the location resolution process that will govern the filtering process performed by word filtering mechanism 102 in accordance with location data to be resolved. The scenarios consider use of any of several different types of inputs by which a location context will be determined.

[0063] Scenario One: Geographic area of interest to be determined by entered zip code and/or city. Zip code and/or city are entered into mobile device 250 by the user. Possible clients used by mobile device 250 to accept the entered zip code and/or city include a voice browser and/or a multimodal client. In the case of a phone call, the client is not the phone itself, but the voice browser (from the server point of view). Location data are provided in the form of a city and/or a zip code. The geographic area of interest corresponding to this location data is determined in the form of a city or zip code. In this case, location resolution is performed locally. The zip code and/or city are entered into mobile device 250 using DTMF keypresses and/or voice. An HTTP request is issued in response to the entered information after the information is received at input mechanism 201. In this case, word subset list 600 is generated by selecting a set of words and/or grammar pertaining to the entered city and/or zip code. The following steps are performed:

[0064] 1. User connects to geographic data acquisition mechanism 101 using mobile device 250. If mobile device 250 represents a multimodal client, an HTTP request is made by the client to a document server of word filtering mechanism 102 (FIGs. 4 and 5). If mobile device 250 represents a wireless telephone, an http request is made by the voice browser ("speaking" with the telephone) to the document server of word filtering mechanism 102.

[0065] 2. A first page is loaded into mobile device 250 or loaded onto the voice browser. The page is prompting the user to enter a zip code. The zip code by either be entered by DTMF (touch tones) or by voice. The grammar used to recognize the zip code is static.

Google Exhibit 1019
Page 00913

[0066] 3. The user enters the zip code into input mechanism 201. A second HTTP request is made by the client, with the zip code value as a parameter. As location-based service 203 is not employed in this scenario to resolve the location, location resolution is "local". The geographic area of interest is evaluated at geographic area of interest defining mechanism 204, and no further refinement of the geographical area of interest takes place.

[0067] 4. A second page is loaded into mobile device 250 (the "reduced" page). The document server of word filtering mechanism 102 computes reduced grammars associated with the page in the form of word subset list 600.

[0068] 5. The user can vocally interact with the page over mobile device 250, as the page is optimized for its location

[0069] Note: The geometry specification refinement mechanism 205 is not used in this scenario. If this mechanism were to be used, a "remote" location resolution must be performed so as to enable sending a map (for example) back to the multimodal client of mobile device 250.

[0070] Scenario Two: Geographic area of interest to be determined automatically using Caller ID or Cell ID.

[0071] Possible clients used by mobile device 250 for Caller Id or Cell ID include a voice browser and/or a multimodal client. Location data are provided in the form of a caller ID or cell ID parameter. The geographic area of interest is determined as a zip code or as an area ID that identifies a geographic area. Location resolution is performed remotely, not locally. Input mechanism 201 is not employed for purposes of location resolution. Mobile device 250 is programmed to issue an HTTP request while a first page is retrieved from a document server of word filtering mechanism 102. In this scenario, words and/or grammar from element database 500 are filtered to prepare word subset list 600 based upon ANI or Cell-ID information. Input mechanism 201 is completely transparent, in the sense that the request that trims the word filtering process appears at the first stage of the interaction, and there is no need of a data collecting stage, as in the previous scenario. The HTTP request can be issued by a voice browser (that provides ANI info) or a multimodal client at mobile device 250.

[0072] The following operational sequence is performed:

[0073] 1. User connects to geographic data acquisition mechanism 101 using mobile device 250. If the user is using a multimodal client at mobile device 250, an HTTP request is made by the client to the document server of word filtering mechanism 102. If mobile device 250 is a wireless telephone, an http request is made by the voice browser ("speaking" with the telephone) to the document server of word filtering mechanism 102. The ANI and/or the Cell ID are passed to the document server as part of the HTTP request. Location-based service 203 determines the geographic area of interest, in this case following the remote path indicated in FIG. 5 which commences at mechanism 202. Location-based service also determines geographic position of the user, an area of uncertainty with respect to the user's position, and one or more associated zip codes.

[0074] 2. A page is downloaded to mobile device 250 or loaded onto the voice browser. The geographic context is used by word filtering mechanism 102 to compute a word subset list 600 associated with the page in accordance with the previously determined geographic area of interest.

[0075] 3. The user can vocally interact with the page using mobile device 250. The page is optimized for the user's geographic area of interest.

[0076] Scenario Three: Geographic area of interest to be determined automatically using GPS data. Possible clients used by mobile device 250 for GPS include a multimodal client. Location data are provided in the form of GPS coordinates. The geographic area of interest is determined as a zip code, as (x,y) or (x,y,z) coordinates, optionally incorporating an area of uncertainty. Location resolution is performed locally, not remotely. Input mechanism 201 is not employed to determine geographic area of interest. An HTTP request is issued by mobile device 250 during the first page fetch. In this case, word filtering mechanism 102 is preparing a word subset list 600 by filtering element database 500 using GPS data. Input mechanism 201 is transparent, in the sense that the HTTP request that results in word filtering appears at the first stage of the interaction. Unlike the first scenario, there is no need for a data collecting stage. The following steps are performed:

[0077] 1. User connects to geographic data acquisition mechanism 101 using mobile device 250. If mobile device 250 utilizes a multimodal client, an HTTP request is made by the client to a document server of word filtering mechanism 102. The GPS information is sent as part of the http request. Location-based service 203 determines the geographic area of interest by adding coarse localization to the GPS information, and the coarse localization is sent to geographic area of interest defining mechanism 204.

[0078] 2. A page is downloaded to mobile device 250: The geographic area of interest in the form of the coarse and/or refined localization received at geographic area of interest defining mechanism 204 is used by word filtering mechanism 102 to compute a word subset list 600 and to place the word subset list into a page that is sent to mobile device 250, using the two-step filtering process described above.

[0079] 3. The user can vocally interact with the page, optimized for the geographic area of interest. Note:

[0080] 1. In cases where geometry specification refinement mechanism 205 is employed, an intermediate page is downloaded to mobile device 250, asking to the user to precise its area of interest (before step 2):

[0081] 1.a - An intermediate (i.e., "localization") page is downloaded to mobile device 250.

[0082] 1.b - The user inputs an area of interest into

Google Exhibit 1019
Page 00914

mobile device 250.

[0083] 1.c - A refined geographic area of interest is sent by mobile device 250 and incorporated into the geographic area of interest defined by geographic area of interest defining mechanism 204.

[0084] In contrast to prior art techniques, the present invention does not employ acoustic models to improve speech recognition accuracy. Instead, the invention reduces the total number of words that need to be considered by a speech recognition algorithm. The number of words is reduced by excluding words that do not relate to a geographic area of interest. In this manner, the reduced processing burden makes it possible to handle speech recognition at the handset. Moreover, the accuracy of the speech recognition algorithm is improved.

[0085] To summarize, the advantages of the invention include more accurate speech recognition due to utilization of a database that has been reduced in size (i.e., word subset list 600). The burden on communication resources is decreased by sending the reduced database to the handset, whereupon the handset performs speech recognition using only the reduced size database instead of a much larger database. The decreased burden on communication resources results in faster, more efficient speech recognition.

[0086] Thus, while there have been shown and described novel features of the invention as applied to embodiments thereof, it will be understood that various omissions and substitutions and changes in the form and details of the devices illustrated, and in their operation, may be made by those skilled in the art without departing from the scope of the invention, which is solely defined by the appended claims. For example, although an embodiment of the invention was described in terms of a user refining an automatically determined geographic area of interest, it is also possible for a user to augment an automatically determined coarse geographic area of interest.

[0087] In view of the foregoing, it is expressly intended that all combinations of those elements and/or method steps which perform substantially the same function in substantially the same way to achieve the same results may be incorporated in any other disclosed or described or suggested form or embodiment as a general matter of design choice in as much as they fall within the scope of the claims appended hereto.

Claims

1. A method for use with a mobile handset equipped to communicate with a remote server over a wireless communications network, the remote server equipped to access an element database (500) including words that describe elements at each of a plurality of geographically defined locations, the method comprising the steps of:

   determining (202) geographic data for the mobile handset following a request; and
   using (102) the determined geographic data to automatically select by the remote server a subset (600) of words for use with a speech recognition procedure;

   wherein the subset of words is selected from the element database, thereby reducing the number of words used in the speech recognition procedure.

2. The method of claim 1 wherein the determined geographic data includes at least one of global positioning system (GPS) position data, cell identity (Cell-ID) data, caller identification (Caller ID) data, place name data, or zip code data.

3. The method of claim 1 wherein the geographic data are derived from a signal received from the mobile handset, said signal being related to a current position of said mobile handset or to a manual input entered into said mobile handset.

4. The method of any preceding claim wherein the elements include at least one of street names, businesses, merchants, points of interest, transportation facilities, individual households, activities, or landmarks.

5. The method of any preceding claim wherein the mobile handset receives the subset of words from the remote server and executes the speech recognition procedure based upon the received subset of words, thereby reducing use of the communications network and the remote server.

6. A speech recognition system comprising:

   a mobile handset equipped to communicate with a remote server over a wireless network,
   a data acquisition mechanism (101) for determining geographic data for the mobile handset following a request input; and
   a selection mechanism (102) for using the determined geographic data to automatically select a subset of words (106) for use with a speech recognition procedure;

   wherein the subset of words is selected by the remote server from an element database (500) said remote server has access to, said element database including words that describe elements at each of a plurality of geographically defined locations, thereby reducing the number of words used in the speech recognition procedure.

7. The speech recognition system of claim 6 wherein the geographic data are derived from a signal re-

Google Exhibit 1019
Page 00915

ceived from the mobile handset, said signal being related to a current position of said mobile handset or to a manual input entered into said mobile handset.

8. The speech recognition system of claim 6 wherein the determined geographic data includes at least one of global positioning system (GPS) position data, cell identity (Cell-ID) data, caller identification (Caller ID) data, place name data, or zip code data.

9. The speech recognition system of any of claims 6 to 8 wherein the elements include at least one of street names, businesses, merchants, points of interest, transportation facilities, individual households, activities, or landmarks.

10. A speech recognition system for associating received speech with words retrieved from an element database, the system comprising:

means for determining a geographic area of interest following a request wherein speech is to be received, said means further comprising:

(i) a portable location determining mechanism associated with a mobile handset for generating an indication signal indicative of current geographic location, and
(ii) a server, in communication with the portable location determining mechanism, programmed to determine a geographic area of interest from the indication signal;

means for selecting a subset of words from the element database based upon the geographic area of interest; said means comprising the server programmed to extract from the element database only words that are associated with the geographic area of interest; and,
means for associating the received speech with words selected from the subset of words, said means being implemented on the mobile handset.

11. The speech recognition system of claim 10 further comprising:

means for accepting a signal from at least one graphical user interface for selecting a portion of the determined geographic area of interest to thereby specify a further limited geographic area of interest; and
means for further reducing the subset of words according to the further limited geographic area of interest to generate a further subset of words, such that the means for associating received speech with words only selects words from the further subset of words.

12. The speech recognition system of claim 11 wherein the means for accepting a signal includes:

an electronic display for displaying a map of the determined geographic area of interest; and
a processing mechanism for combining the accepted signal with the map of the determined geographic area of interest so as to cause a display of the further limited geographic area of interest on the electronic display.

13. A mobile handset (250) comprising speech recognition means for association of received speech with words received from a remote database, the mobile handset comprising:

means for receiving a request;
means for providing a signal representing a geographic area of interest following said request;
means for transmitting said signal to a remote server and for receiving from the remote server a subset of words from the remote database based upon the geographic area of interest; and
means for associating the received speech with words selected from the subset of words.

14. The mobile handset of claim 13 wherein the means for providing a signal comprises means for determining a geographic area of interest wherein speech is to be received.

15. The mobile handset of claim 13 or 14 further comprising:

means (201) for accepting at least one input used to determine a refined geographic area of interest in the geographic area of interest;
means (205) for reducing the subset of words according to the refined geographic area of interest;

and wherein:

the means for associating the received speech is arranged for association of speech only with the reduced subset of words.

Patentansprüche

1. Verfahren zur Verwendung mit einem mobilen Handapparat, der dafür ausgestattet ist, über ein drahtloses Kommunikationsnetz mit einem abgesetzten Server zu kommunizieren, wobei der abgesetzte Server dafür ausgestattet ist, auf eine Elementdatenbank (500) zuzugreifen, die Wörter enthält, die

Google Exhibit 1019
Page 00916

Elemente an jedem von mehreren geographisch definierten Orten beschreiben, wobei das Verfahren die folgenden Schritte umfaßt:

Bestimmen (202) geographischer Daten für den mobilen Handapparat nach einer Anforderung und Verwenden (102) der bestimmten geographischen Daten zum automatischen Auswählen einer Teilmenge (600) von Wörtern zur Verwendung mit einer Spracherkennungsprozedur durch den abgesetzten Server;

wobei die Teilmenge von Wörtern aus der Elementdatenbank ausgewählt wird, wodurch die Anzahl der in der Spracherkennungsprozedur verwendeten Wörter reduziert wird.

2. Verfahren nach Anspruch 1, wobei die bestimmten geographischen Daten Positionsdaten des globalen Navigationssystems (GPS) und/oder Daten der Zellenidentität (Cell-ID) und/oder Daten der Anruferidentifikation (Caller ID) und/oder Städtenamendaten und/oder Postleitzahldaten umfassen.

3. Verfahren nach Anspruch 1, wobei die geographischen Daten aus einem aus dem mobilen Handapparat empfangenen Signal abgeleitet werden, wobei das Signal eine aktuelle Position des mobilen Handapparats oder eine in den mobilen Handapparat eingegebene manuelle Eingabe betrifft.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei die Elemente Straßennamen und/oder Unternehmen und/oder Händler und/oder interessierende Punkte und/oder Transporteinrichtungen und/oder individuelle Haushalte und/oder Aktivitäten und/oder Sehenswürdigkeiten umfassen.

5. Verfahren nach einem der vorhergehenden Ansprüche, wobei der mobile Handapparat die Teilmenge von Wörtern von dem abgesetzten Server empfängt und auf der Basis der empfangenen Teilmenge von Wörtern die Spracherkennungsprozedur ausführt und **dadurch** die Benutzung des Kommunikationsnetzes und des abgesetzten Servers reduziert.

6. Spracherkennungssystem, umfassend:

einen mobilen Handapparat, der dafür ausgestattet ist, über ein drahtloses Netz mit einem abgesetzten Server zu kommunizieren,
einen Datenbeschaffungsmechanismus (101) zum Bestimmen von geographischen Daten für den mobilen Handapparat nach einer Anforderungseingabe und
einen Auswahlmechanismus (102) zum Verwenden der bestimmten geographischen Daten zum automatischen Auswählen einer Teilmenge von Wörtern (106) zur Verwendung mit einer Spracherkennungsprozedur;

wobei die Teilmenge von Wörtern durch den abgesetzten Server aus einer Elementdatenbank (500) ausgewählt wird, zu der der abgesetzte Server Zugang hat, wobei die Elementdatenbank Wörter enthält, die Elemente an jedem von mehreren geographisch definierten Orten beschreiben, wodurch die Anzahl der bei der Spracherkennungsprozedur verwendeten Wörter reduziert wird.

7. Spracherkennungssystem nach Anspruch 6, wobei die geographischen Daten aus einem aus dem mobilen Handapparat empfangenen Signal abgeleitet werden, wobei das Signal eine aktuelle Position des mobilen Handapparats oder eine in den mobilen Handapparat eingegebene manuelle Eingabe betrifft.

8. Spracherkennungssystem nach Anspruch 6, wobei die bestimmten geographischen Daten Positionsdaten des globalen Navigationssystems (GPS) und/oder Daten der Zellenidentität (Cell-ID) und/oder Daten der Anruferidentifikation (Caller ID) und/oder Städtenamendaten und/oder Postleitzahldaten umfassen.

9. Spracherkennungssystem nach einem der Ansprüche 6 bis 8, wobei die Elemente Straßennamen und/oder Unternehmen und/oder Händler und/oder interessierende Punkte und/oder Transporteinrichtungen und/oder individuelle Haushalte und/oder Aktivitäten und/oder Sehenswürdigkeiten umfassen.

10. Spracherkennungssystem zum Assoziieren von empfangener Sprache mit aus einer Elementdatenbank abgerufenen Wörtern, wobei das System folgendes umfaßt:

Mittel zum Bestimmen eines interessierenden geographischen Gebiets nach einer Aufforderung, wobei Sprache zu empfangen ist, wobei die Mittel ferner folgendes umfassen:

i) einen mit einem mobilen Handapparat assoziierten tragbaren Ortsbestimmungsmechanismus zum Erzeugen eines Anzeigesignals, das einen aktuellen geographischen Ort anzeigt und
ii) einen mit dem tragbaren Ortsbestimmungsmechanismus kommunizierenden Server, der dafür programmiert ist, aus dem Anzeigesignal ein interessierendes geographisches Gebiet zu bestimmen;

Mittel zum Auswählen einer Teilmenge von Wörtern aus der Elementdatenbank auf der Ba-

Google Exhibit 1019
Page 00917

sis des interessierenden geographischen Gebiets; wobei die Mittel den Server umfassen, der dafür programmiert ist, aus der Elementdatenbank nur Wörter zu extrahieren, die mit dem interessierenden geographischen Gebiet assoziiert sind und
Mittel zum Assoziieren der empfangenen Sprache mit aus der Teilmenge von Wörtern ausgewählten Wörtern, wobei die Mittel auf dem mobilen Handapparat implementiert sind.

11. Spracherkennungssystem nach Anspruch 10, ferner umfassend:

Mittel zum Annehmen eines Signals von mindestens einer graphischen Benutzeroberfläche zum Auswählen eines Teils des bestimmten interessierenden geographischen Gebiets, um **dadurch** ein weiter eingeschränktes interessierendes geographisches Gebiet zu spezifizieren und
Mittel zum weiteren Reduzieren der Teilmenge von Wörtern gemäß dem weiter eingeschränkten interessierenden geographischen Gebiet, um eine weitere Teilmenge von Wörtern zu erzeugen, dergestalt, daß die Mittel zum Assoziieren empfangener Sprache mit Wörtern nur Wörter aus der weiteren Teilmenge von Wörtern auswählen.

12. Spracherkennungssystem nach Anspruch 11, wobei die Mittel zum Annehmen eines Signals folgendes umfassen:

ein elektronisches Display zum Anzeigen einer Karte des bestimmten interessierenden geographischen Gebiets und
einen Verarbeitungsmechanismus zum Kombinieren des angenommenen Signals mit der Karte des bestimmten interessierenden geographischen Gebiets, um so eine Anzeige des weiter eingeschränkten interessierenden geographischen Gebiets auf dem elektronischen Display zu bewirken.

13. Mobiler Handapparat (150) mit Spracherkennungsmitteln zum Assoziieren empfangener Sprache mit aus einer abgesetzten Datenbank empfangenen Wörtern, wobei der mobile Handapparat folgendes umfaßt:

Mittel zum Empfangen einer Anforderung;
Mittel zum Bereitstellen eines Signals, das ein interessierendes geographisches Gebiet repräsentiert, nach der Anforderung;
Mittel zum Senden des Signals zu einem abgesetzten Server und zum Empfangen einer Teilmenge von Wörtern aus der abgesetzten Datenbank von dem abgesetzten Server auf der Basis des interessierenden geographischen Gebiets und
Mittel zum Assoziieren der empfangenen Sprache mit aus der Teilmenge von Wörtern ausgewählten Wörtern.

14. Mobiler Handapparat nach Anspruch 13, wobei die Mittel zum Bereitstellen eines Signals Mittel zum Bestimmen eines interessierenden geographischen Gebiets umfaßt, worin Sprache empfangen werden soll.

15. Mobiler Handapparat nach Anspruch 13 oder 14, ferner umfassen:

Mittel (210) zum Annehmen mindestens einer Eingabe, mit der ein verfeinertes interessierendes geographisches Gebiet in dem interessierenden geographischen Gebiet bestimmt wird;
Mittel (205) zum Reduzieren der Teilmenge von Wörtern gemäß dem verfeinerten interessierenden geographischen Gebiet und wobei

die Mittel zum Assoziieren der empfangenen Sprache dafür ausgelegt sind, Sprache nur mit der reduzierten Teilmenge von Wörtern zu assoziieren.

Revendications

1. Procédé pour une utilisation avec un combiné mobile équipé pour communiquer avec un serveur distant sur un réseau de communication sans fil, le serveur distant étant équipé pour accéder à une base de données d'éléments (500) comprenant des mots qui décrivent des éléments pour chacun d'une pluralité d'emplacements définis géographiquement, le procédé comprenant les étapes de :

détermination (202) de données géographiques pour le combiné mobile suite à une demande ; et utilisation (102) des données géographiques déterminées pour sélectionner automatiquement par le serveur distant un sous-ensemble (600) de mots pour une utilisation avec une procédure de reconnaissance vocale ;

dans lequel le sous-ensemble de mots est sélectionné parmi la base de données d'éléments, réduisant de ce fait le nombre de mots utilisés dans la procédure de reconnaissance vocale.

2. Procédé selon la revendication 1 dans lequel les données géographiques déterminées comprennent au moins une parmi des données de position du système de positionnement mondial (GPS), des données d'identité de cellule (identifiant de cellule), des

Google Exhibit 1019
Page 00918

données d'identification d'appelant (identifiant d'appelant), des données de nom de lieu ou des données de code postal.

3. Procédé selon la revendication 1 dans lequel les données géographiques sont déduites d'un signal reçu du combiné mobile, ledit signal étant lié à une position actuelle dudit combiné mobile ou à une entrée manuelle entrée dans ledit combiné mobile.

4. Procédé selon une revendication précédente quelconque dans lequel les éléments comprennent au moins un parmi des noms de rues, des entreprises, des commerçants, des centres d'intérêt, des installations de transport, des maisons individuelles, des activités ou des points de repère.

5. Procédé selon une revendication précédente quelconque dans lequel le combiné mobile reçoit le sous-ensemble de mots du serveur distant et exécute la procédure de reconnaissance vocale sur la base du sous-ensemble reçu de mots, réduisant de ce fait l'utilisation du réseau de communication et du serveur distant.

6. Système de reconnaissance vocale comprenant :

un combiné mobile équipé pour communiquer avec un serveur distant sur un réseau sans fil ;
un mécanisme d'acquisition de données (101) pour déterminer des données géographiques pour le combiné mobile suite à une entrée de demande ; et
un mécanisme de sélection (102) pour utiliser les données géographiques déterminées afin de sélectionner automatiquement un sous-ensemble de mots (106) pour une utilisation avec une procédure de reconnaissance vocale ;

dans lequel le sous-ensemble de mots est sélectionné par le serveur distant à partir d'une base de données d'éléments (500) à laquelle ledit serveur distant a accès, ladite base de données d'éléments comprenant des mots qui décrivent des éléments pour chacun d'une pluralité d'emplacements définis géographiquement, réduisant de ce fait le nombre de mots utilisés dans la procédure de reconnaissance vocale.

7. Système de reconnaissance vocale selon la revendication 6 dans lequel les données géographiques sont déduites d'un signal reçu du combiné mobile, ledit signal étant lié à une position actuelle dudit combiné mobile ou à une entrée manuelle entrée dans ledit combiné mobile.

8. Système de reconnaissance vocale selon la revendication 6 dans lequel les données géographiques déterminées comprennent au moins une parmi des données de position du système de positionnement mondial (GPS), des données d'identité de cellule (identifiant de cellule), des données d'identification d'appelant (identifiant d'appelant), des données de nom de lieu ou des données de code postal.

9. Système de reconnaissance vocale selon l'une quelconque des revendications 6 à 8 dans lequel les éléments comprennent au moins un parmi des noms de rues, des entreprises, des commerçants, des centres d'intérêt, des installations de transport, des maisons individuelles, des activités ou des points de repère.

10. Système de reconnaissance vocale pour associer une voix reçue à des mots récupérés d'une base de données d'éléments, le système comprenant :

des moyens pour déterminer une zone géographique d'intérêt suite à une demande dans laquelle une voix doit être reçue, lesdits moyens comprenant en outre :

(i) un mécanisme de détermination d'emplacement de portable associé à un combiné mobile pour générer un signal d'indication révélateur d'un emplacement géographique actuel, et
(ii) un serveur, en communication avec le mécanisme de détermination d'emplacement de portable, programmé pour déterminer une zone géographique d'intérêt à partir du signal d'indication ;

des moyens pour sélectionner un sous-ensemble de mots parmi la base de données d'éléments sur la base de la zone géographique d'intérêt ; lesdits moyens comprenant le serveur programmé pour extraire de la base de données d'éléments seulement des mots qui sont associés à la zone géographique d'intérêt ; et
des moyens pour associer la voix reçue à des mots sélectionnés parmi le sous-ensemble de mots, lesdits moyens étant mis en oeuvre sur le combiné mobile.

11. Système de reconnaissance vocale selon la revendication 10 comprenant en outre :

des moyens pour accepter un signal d'au moins une interface utilisateur graphique pour sélectionner une partie de la zone géographique d'intérêt déterminée afin de spécifier de ce fait une zone géographique d'intérêt limitée supplémentaire ; et
des moyens pour réduire en outre le sous-ensemble de mots selon la zone géographique

Google Exhibit 1019
Page 00919

d'intérêt limitée supplémentaire afin de générer un sous-ensemble supplémentaire de mots, de telle sorte que les moyens pour associer une voix reçue à des mots sélectionnent seulement des mots parmi le sous-ensemble supplémentaire de mots.

12. Système de reconnaissance vocale selon la revendication 11 dans lequel les moyens pour accepter un signal comprennent :

un affichage électronique pour afficher une carte de la zone géographique d'intérêt déterminée ; et
un mécanisme de traitement pour combiner le signal accepté avec la carte de la zone géographique d'intérêt déterminée de façon à provoquer un affichage de la zone géographique d'intérêt limitée supplémentaire sur l'affichage électronique.

13. Combiné mobile (250) comprenant des moyens de reconnaissance vocale pour une association d'une voix reçue à des mots reçus d'une base de données distante, le combiné mobile comprenant :

des moyens pour recevoir une demande ;
des moyens pour fournir un signal représentant une zone géographique d'intérêt suite à ladite demande ;
des moyens pour transmettre ledit signal à un serveur distant et recevoir du serveur distant un sous-ensemble de mots provenant de la base de données distante sur la base de la zone géographique d'intérêt ; et
des moyens pour associer la voix reçue à des mots sélectionnés parmi le sous-ensemble de mots.

14. Combiné mobile selon la revendication 13 dans lequel les moyens pour fournir un signal comprennent des moyens pour déterminer une zone géographique d'intérêt dans laquelle une voix doit être reçue.

15. Combiné mobile selon la revendication 13 ou 14 comprenant en outre:

des moyens (201) pour accepter au moins une entrée utilisée pour déterminer une zone géographique affinée d'intérêt dans la zone géographique d'intérêt ;
des moyens (205) pour réduire le sous-ensemble de mots selon la zone géographique affinée d'intérêt;

et dans lequel :

les moyens pour associer la voix reçue sont agencés pour une association d'une voix seulement avec le sous-ensemble réduit de mots.

Google Exhibit 1019
Page 00920

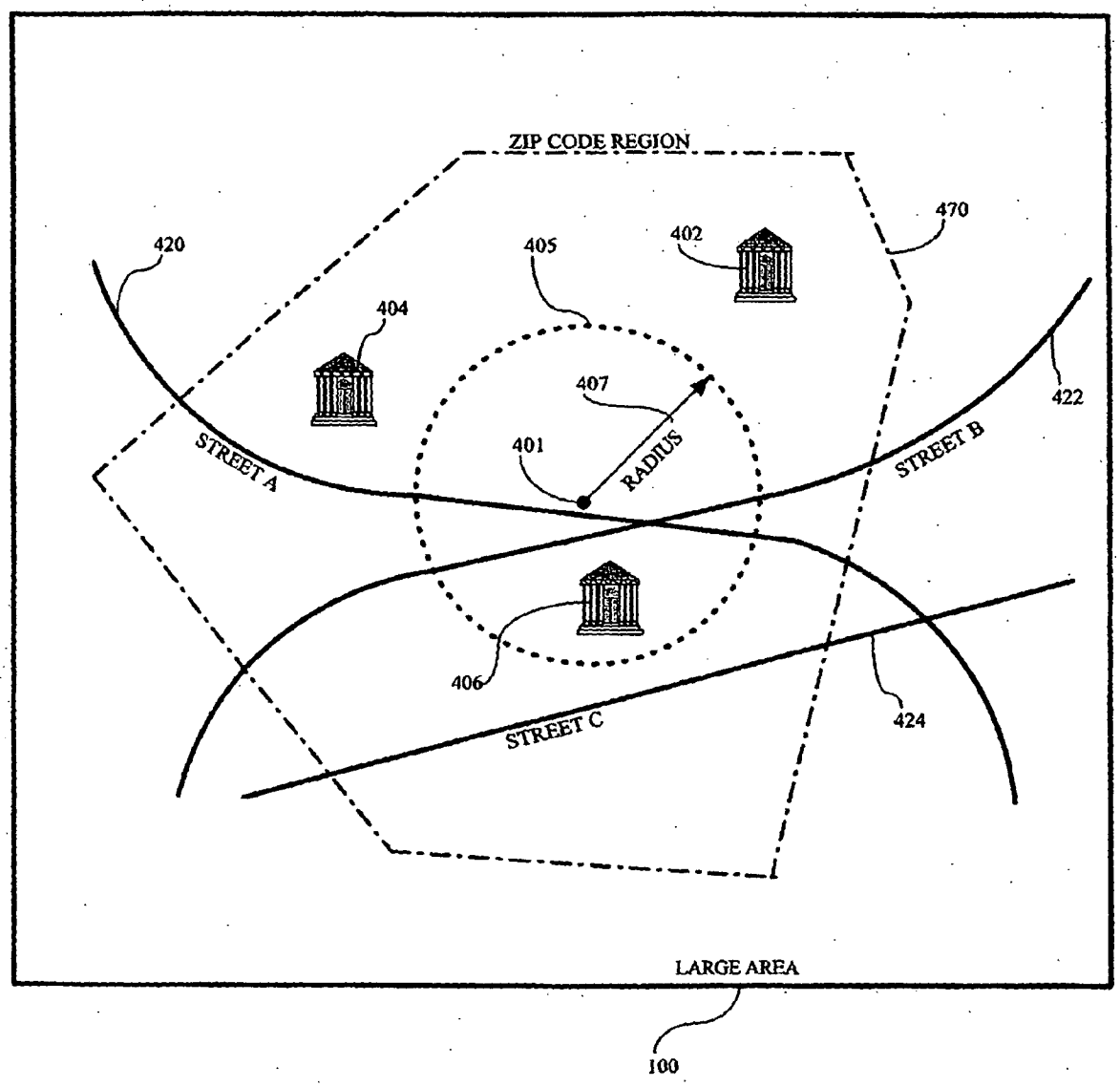


FIG. 1

Google Exhibit 1019
Page 00921

500

ELEMENT DATABASE

ELEMENT NAME FIELD 501	PARAMETER FIELD FOR THE REFINED GEOGRAPHIC AREA OF INTEREST 503	PARAMETER FIELD FOR THE COARSE GEOGRAPHIC AREA OF INTEREST 505
FIRST BUILDING	{[603001, 2428833]}	10017
SECOND BUILDING	{[602298, 2428830]}	10017
THIRD BUILDING	{[603010, 2428836]}	10017
STREET A	{[603012, 2428836],[603013, 24428836]}	10017, 10036
STREET B	{[603014, 2428836],[603015, 24428836]}	10017, 10055
STREET C	{[679010, 2428836],[680000, 24428836]}	10017, 12533
...	...	...
DEVON FARMS ROAD	{[673010, 2428836],[675000, 24428836]}	12582

FIG. 2

600

WORD SUBSET LIST

ELEMENT NAME FIELD 601
FIRST BUILDING
STREET A
STREET B

FIG. 3

Google Exhibit 1019
Page 00922

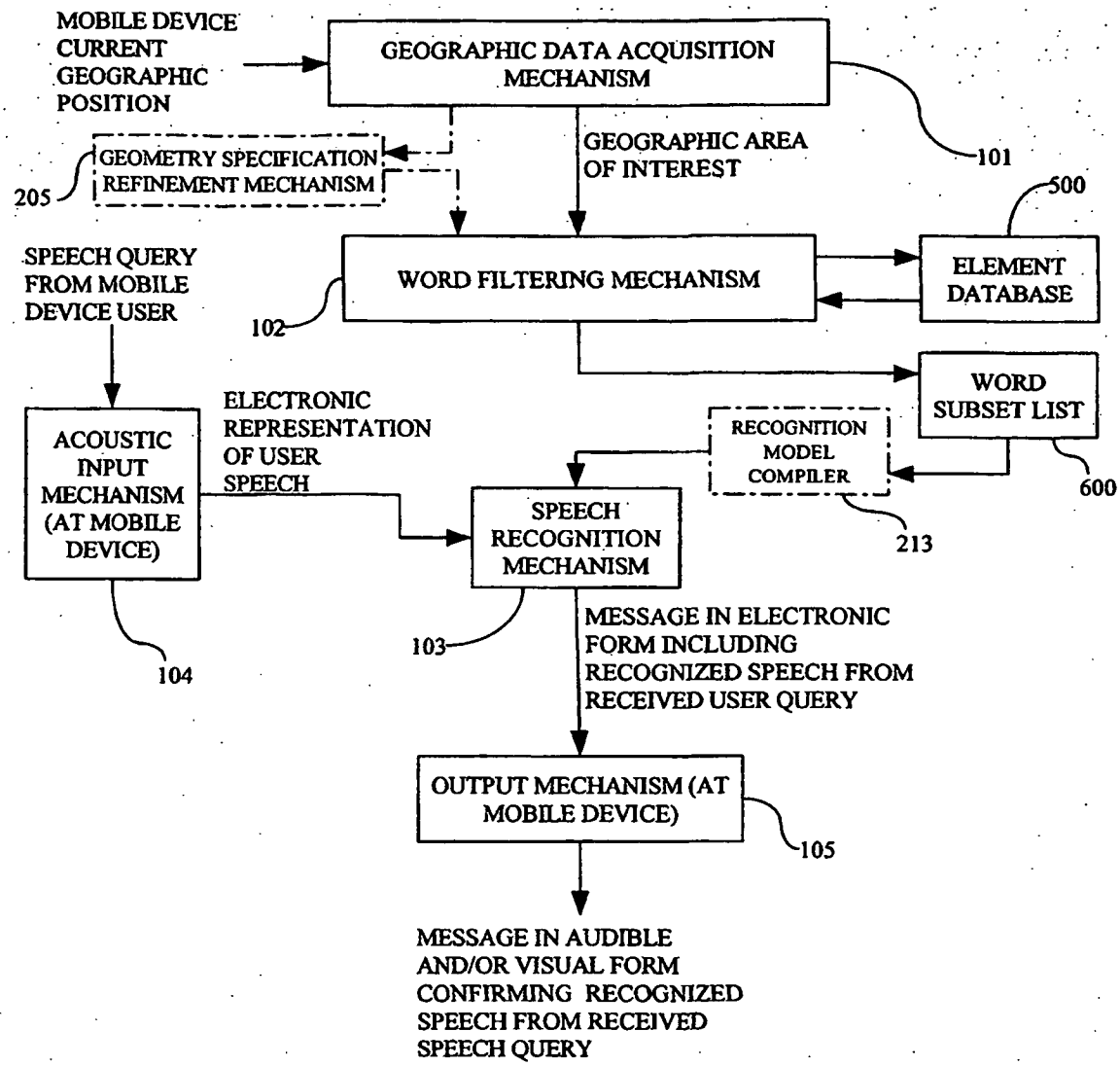


FIG. 4

Google Exhibit 1019
Page 00923

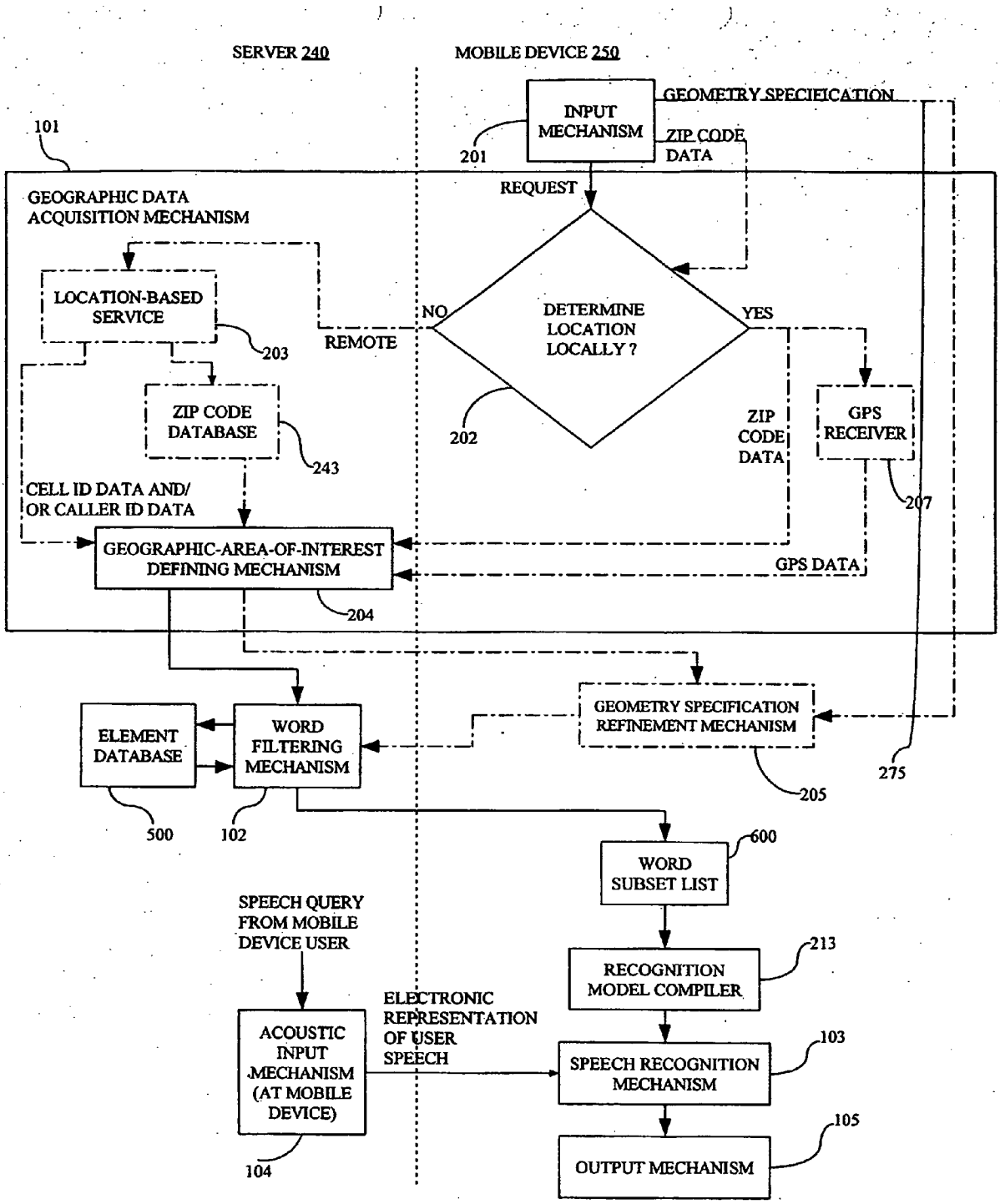


FIG. 5

Google Exhibit 1019
Page 00924

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 2003236099 A **[0007] [0009] [0010]**
- EP 1233407 A **[0011]**
- US 6230132 B **[0012]**
- EP 0661688 A **[0013]**

Google Exhibit 1019
Page 00925

Electronic Patent Application Fee Transmittal

Application Number:	12705311
Filing Date:	12-Feb-2010
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Filer:	Chen Li/Jennifer Hake
Attorney Docket Number:	090611-0384800

Filed as Large Entity

Utility under 35 USC 111(a) Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Extension - 2 months with $0 paid	1252	1	560	560

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
		Total in USD ($)		**740**

Electronic Acknowledgement Receipt

EFS ID:	12723295
Application Number:	12705311
International Application Number:	
Confirmation Number:	8125
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Customer Number:	909
Filer:	Chen Li/Jennifer Hake
Filer Authorized By:	Chen Li
Attorney Docket Number:	090611-0384800
Receipt Date:	07-MAY-2012
Filing Date:	12-FEB-2010
Time Stamp:	20:07:34
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	$740
RAM confirmation Number	6616
Deposit Account	033975
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.20 (Post Issuance fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Extension of Time	090611_0384800_Extension.pdf	88152 274585c9e249ebafcb7552453ec041f0eb2a6c2d	no	1
Warnings:					
Information:					
2		090611_0384800_Response.pdf	889945 f527c703efbb067b310a479870fade777e0d3189	yes	16

Multipart Description/PDF files in .zip description

Document Description	Start	End
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Claims	2	12
Applicant Arguments/Remarks Made in an Amendment	13	16

Warnings:					
Information:					
3	Transmittal Letter	0384800_IDSTransmittal.pdf	51159 ab6d2a8872e0d31e6f88f9c167c156256d4abb23	no	2
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	0384800_IDSPTO1449.pdf	63392 70e43b544037e56bc59c166677186c79f7adb72e	no	6
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
5	Foreign Reference	EP1320043A2.pdf	1414854 ff924e87b0c0a160e04e60fc68d69fc8493c317e	no	28
Warnings:					
Information:					
6	Foreign Reference	EP1646037.pdf	266295 8d9c9ea22bd5808311bb8a8d8d653ec96d2c2f06	no	20

Warnings:					
Information:					
7	Non Patent Literature	Statement_in_Accordance_with_EPO_Notice.pdf	86078 aa691f8e43b79b60df22073f977c16f6d0354668	no	1
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	32610 7dddeff61819a9ac4b11e9978dc8259b203ea2e8	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	2892485	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/22 (09-11)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARMENT OF COMMERCE
Under the paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless if displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) 090611-0384800
Application Number 12/705,311	Filed February 12, 2010
For MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE	
Art Unit 2626	Examiner Fariba Sirjani

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	
☐ One month (37 CFR 1.17(a)(1))	$150	$75	$
☒ Two months (37 CFR 1.17(a)(2))	$560	$280	$ 560.00
☐ Three months (37 CFR 1.17(a)(3))	$1270	$635	$
☐ Four months (37 CFR 1.17(a)(4))	$1980	$990	$
☐ Five months (37 CFR 1.17(a)(5))	$2690	$1345	$

☐ Applicant claims small entity status. See 37 CFR 1.27.

☐ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Director has already been authorized to charge fees in this application to a Deposit Account.

☒ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 033975.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the ☐ applicant/inventor.

☐ assignee of record of the entire interest. See 37 CFR 3.71. Statement under 37 CFR 3.73(b) is enclosed (Form PTO/SB/96).

☒ attorney or agent of record. Registration Number 67448

☐ attorney or agent under 37 CFR 1.34.
Registration number if acting under 37 CFR 1.34 _____.

/Chen Li/ — Signature
May 7, 2012 — Date
Chen Li — Typed or printed name
(703) 770-7900 — Telephone Number

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below.

☒ Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.136(a). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 6 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

American LegalNet, Inc.
www.FormsWorkFlow.com

Electronic Patent Application Fee Transmittal

Application Number:	12705311
Filing Date:	12-Feb-2010
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Filer:	Chen Li/Jennifer Hake
Attorney Docket Number:	090611-0384800

Filed as Small Entity

Utility under 35 USC 111(a) Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Claims in excess of 20	2202	14	30	420
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
			Total in USD ($)	**420**

Electronic Acknowledgement Receipt

EFS ID:	12723417
Application Number:	12705311
International Application Number:	
Confirmation Number:	8125
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Customer Number:	909
Filer:	Chen Li/Jennifer Hake
Filer Authorized By:	Chen Li
Attorney Docket Number:	090611-0384800
Receipt Date:	07-MAY-2012
Filing Date:	12-FEB-2010
Time Stamp:	20:31:11
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	$420
RAM confirmation Number	6754
Deposit Account	033975
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.20 (Post Issuance fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Fee Worksheet (SB06)	fee-info.pdf	30591 7a2d1149a7d7bd41b53dc93f626a8df9dae6d3f7	no	2

Warnings:

Information:

Total Files Size (in bytes):	30591

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/06 (07-06)
Approved for use through 1/31/2007. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number	Filing Date	☐ To be Mailed
12/705,311	02/12/2010	

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY ☒ RATE ($)	SMALL ENTITY ☒ FEE ($)	OR	OTHER THAN SMALL ENTITY RATE ($)	OTHER THAN SMALL ENTITY FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A			N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =		OR	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =			X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							
			TOTAL			TOTAL	

* If the difference in column 1 is less than zero, enter "0" in column 2.

APPLICATION AS AMENDED – PART II

AMENDMENT 05/07/2012	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE ($)	SMALL ENTITY ADDITIONAL FEE ($)	OR	OTHER THAN SMALL ENTITY RATE ($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 44	Minus	** 30	= 14	X $30 =	420	OR	X $ =	
Independent (37 CFR 1.16(h))	* 2	Minus	***3	= 0	X $125 =	0	OR	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))									
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
					TOTAL ADD'L FEE	420	OR	TOTAL ADD'L FEE	

AMENDMENT	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)	OR	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =		OR	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =		OR	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))									
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
					TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

Legal Instrument Examiner:
/STELLA LITTLE/

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Attorney's Docket No. 090611-0384800 **PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS :	Robert A. KENNEWICK *et al.*	CONFIRMATION NO.:	8125
SERIAL NUMBER :	12/705,311	EXAMINER:	SIRJANI, FARIBA
FILING DATE :	February 12, 2010	ART UNIT:	2626
FOR :	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE		

REQUEST FOR REFUND

Mail Stop 16
Director of the U.S. Patent and Trademark Office
P. O. Box 1450
Alexandria, VA 22313-1450

Dear Sir/Madam:

I. REFUND REQUEST

Pursuant to 37 CFR 1.26, Applicant was charged $560.00 on May 7, 2012, for a two-month Extension of Time as a large entity. However, Applicant is a small entity and and the charge should have been $280.00. Applicant hereby requests a refund in the amount of $280.00 to be credited to Deposit Account No. 03-3975, in the above-identified application.

II. FEES CHARGED FOR WHICH REFUND REQUESTED

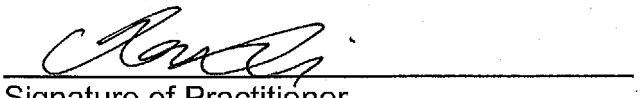
AMOUNT OF REFUND REQUESTED
$280.00

The fee for a two-month extension of time (large entity) was charged to Deposit Account No. 03-3975, when in fact a two-month extension of time (small entity) was necessary.

III. MANNER OF REFUND

Please make refund by crediting Account No. 03-3975.

Respectfully Submitted,



Signature of Practitioner

Date: May 15, 2012

Reg. No.: 67448
Tel. No.: 703.770.7741
Customer No.: 00909

Chen Li
PILLSBURY WINTHROP SHAW PITTMAN LLP
1650 Tysons Boulevard, Suite 1400
McLean, VA 22102-4856

05/08	6448	13465628	524465-0407677	2201	$125.00	$512,600.70
05/08	6911	11269638	019287-0316611	1801	$930.00	$511,670.70
05/08	6909	11269638	019287-0316611	1202	$240.00	$511,430.70
05/08	6910	11269638	019287-0316611	1251	$150.00	$511,280.70
05/08	7194	PCT/US12/36785	34921-407679	1601	$240.00	$511,040.70
05/08	7195	PCT/US12/36785	34921-407679	1701	$1,344.00	$509,696.70
05/08	7196	PCT/US12/36785	34921-407679	1703	$928.00	$508,768.70
05/08	7197	PCT/US12/36785	34921-407679	1714	$415.00	$508,353.70
05/08	7337	3013102	035062-0398127-43381	7205	$100.00	$508,253.70
05/08	7338	3013102	035062-0398127-43381	7208	$200.00	$508,053.70
05/08	7339	3013102	035062-0398127-43381	7206	$100.00	$507,953.70
05/08	8135	13401560	081468-0403261	1111	$620.00	$507,333.70
05/08	8134	13401560	081468-0403261	1011	$380.00	$506,953.70
05/08	8137	13401560	081468-0403261	1051	$130.00	$506,823.70
05/08	8136	13401560	081468-0403261	1311	$250.00	$506,573.70
05/08	9234	10747612	201818-0307164	1403	$1,240.00	$505,333.70
05/08	10044	US1231672	085133-0405489	8021	$40.00	$505,293.70
05/08	10981	61643819	017499-0403344 016719L2	1005	$250.00	$505,043.70
05/08	11214	13396576	048285-0403100 CTV-006-C	2051	$65.00	$504,978.70
05/08	11350	12705311	090611-0384800	1806	$180.00	$504,798.70
05/08	11349	12705311	090611-0384800	1252	$560.00	$504,238.70
05/08	11591	12705311	090611-0384800	2202	$420.00	$503,818.70
05/08	12744	85133606	034888-0396990/RBB	7004	$450.00	$503,368.70
05/08	17336	13213604	052775-0404462	8007	$20.00	$503,348.70
05/09	453	13466355	007277-0407006	2111	$310.00	$503,038.70
05/09	452	13466355	007277-0407006	4011	$95.00	$502,943.70
05/09	454	13466355	007277-0407006	2311	$125.00	$502,818.70
05/09	949	12301348	012140-0376068	1801	$930.00	$501,888.70
05/09	1170	13466403	007277-0407005	4011	$95.00	$501,793.70
05/09	1172	13466403	007277-0407005	2311	$125.00	$501,668.70
05/09	1171	13466403	007277-0407005	2111	$310.00	$501,358.70
05/09	1315	13363340	091230-0403102	8021	$40.00	$501,318.70
05/09	3495	11558795	019287-0317685	8021	$40.00	$501,278.70
05/09	3797	11558795	019287-0317685	8021	$40.00	$501,238.70
05/09	4145	12031984	IMM288 (I103 1280.1)	1202	$420.00	$500,818.70
05/09	4212	PCT/US11/43235	201818-0396560	1705	$219.00	$500,599.70
05/09	4211	PCT/US11/43235	201818-0396560	1605	$600.00	$499,999.70
05/09	4697	10887750	019287-0319552	8021	$40.00	$499,959.70
05/09	4699	13434332	091230-0405189	8021	$40.00	$499,919.70
05/09	5483	13434318	091230-0393065	8021	$40.00	$499,879.70
05/09	5823	11192400	081468-0316985	1504	$300.00	$499,579.70
05/09	5822	11192400	081468-0316985	1501	$1,740.00	$497,839.70
05/09	5824	11192400	081468-0316985	8001	$9.00	$497,830.70
05/09	6362	61586370	091230-0403055	8021	$40.00	$497,790.70
05/09	6753	2570354	081103-0248220	7205	$100.00	$497,690.70
05/09	6754	2570354	081103-0248220	7201	$400.00	$497,290.70
05/09	7805	11352294	201818-0319724	1253	$1,270.00	$496,020.70
05/09	8175	29397639	081627-0395817	1502	$990.00	$495,030.70
05/09	11801	61604246	091230-0404211	1052	$50.00	$494,980.70
05/09	11837	13363340	091230-0403102	1111	$620.00	$494,360.70
05/09	11839	13363340	091230-0403102	1051	$130.00	$494,230.70
05/09	11838	13363340	091230-0403102	1311	$250.00	$493,980.70
05/09	11836	13363340	091230-0403102	1011	$380.00	$493,600.70
05/09	13618	11770945	043395-0378088	1501	$1,740.00	$491,860.70

05/09

Electronic Acknowledgement Receipt

EFS ID:	12782185
Application Number:	12705311
International Application Number:	
Confirmation Number:	8125
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Customer Number:	909
Filer:	Chen Li/Jennifer Hake
Filer Authorized By:	Chen Li
Attorney Docket Number:	090611-0384800
Receipt Date:	15-MAY-2012
Filing Date:	12-FEB-2010
Time Stamp:	14:55:16
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Refund Request	090611_0384800_Request_for_Refund.pdf	155285 cfbd96e2261353bc6ba521a624a878990d83c3c9	no	2

Warnings:	
Information:	

Google Exhibit 1019

Total Files Size (in bytes):	155285

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney's Docket No. 090611-0384800 **PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS :	Robert A. KENNEWICK *et al.*	CONFIRMATION NO.:	8125
SERIAL NUMBER :	12/705,311	EXAMINER:	SIRJANI, FARIBA
FILING DATE :	February 12, 2010	ART UNIT:	2626
FOR :	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE		

REQUEST FOR REFUND

2012 MAY 15 AM 11:22 USPTO

Mail Stop 16
Director of the U.S. Patent and Trademark Office
P. O. Box 1450
Alexandria, VA 22313-1450

Dear Sir/Madam:

I. REFUND REQUEST

Pursuant to 37 CFR 1.26, Applicant was charged $560.00 on May 7, 2012, for a two-month Extension of Time as a large entity. However, Applicant is a small entity and and the charge should have been $280.00. Applicant hereby requests a refund in the amount of $280.00 to be credited to Deposit Account No. 03-3975, in the above-identified application.

II. FEES CHARGED FOR WHICH REFUND REQUESTED

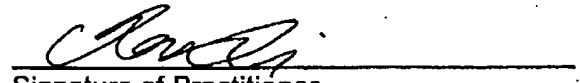
	AMOUNT OF REFUND REQUESTED
	$280.00

The fee for a two-month extension of time (large entity) was charged to Deposit Account No. 03-3975, when in fact a two-month extension of time (small entity) was necessary.

III. MANNER OF REFUND

Please make refund by crediting Account No. 03-3975.

Respectfully Submitted,



Signature of Practitioner

Date: May 15, 2012

Reg. No.: 67448
Tel. No.: 703.770.7741
Customer No.: 00909

Chen Li
PILLSBURY WINTHROP SHAW PITTMAN LLP
1650 Tysons Boulevard, Suite 1400
McLean, VA 22102-4856

Google Exhibit 1019
Page 00941

Electronic Patent Application Fee Transmittal

Application Number:	12705311
Filing Date:	12-Feb-2010
Title of Invention:	MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE
First Named Inventor/Applicant Name:	Robert A. Kennewick
Filer:	Chen Li/Jennifer Hake
Attorney Docket Number:	090611-0384800

Filed as Large Entity

Utility under 35 USC 111(a) Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Extension - 2 months with $0 paid	1252	1	560	560

Adjustment date: 05/18/2012 EEKUBAY1 12705311
05/08/2012 INTEFSW 00006616 033975
01 FC:1252 560.00 CR

05/18/2012 EEKUBAY1 00000012 033975 12705311
01 FC:2252 560 280.00 DA

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
		Total in USD ($)		**740**

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
12/705,311	02/12/2010	Robert A. Kennewick	090611-0384800	8125

909 7590 08/03/2012

Pillsbury Winthrop Shaw Pittman, LLP (NV)
PO Box 10500
McLean, VA 22102

EXAMINER	
SIRJANI, FARIBA	

ART UNIT	PAPER NUMBER
2626	

NOTIFICATION DATE	DELIVERY MODE
08/03/2012	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket_ip@pillsburylaw.com
margaret.drosos@pillsburylaw.com

Office Action Summary	Application No.	Applicant(s)	
	12/705,311	KENNEWICK ET AL.	
	Examiner	Art Unit	
	FARIBA SIRJANI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *07 May 2012*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

5)☒ Claim(s) *56-99* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *56-99* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on _____ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

12)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 56-99 are pending. Claims 56 and 71 are independent. Claims 86-99 are new. New Claims 86-98 depend from Claim 56 and new Claim 99 depends from Claim 71.

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, THIS ACTION IS MADE FINAL.

Response to Arguments

2. Applicant has made the following arguments that hinge on the amendments to the independent Claims 56 and 71:

> Independent claim 56, for example, recites a mobile system for processing natural language utterances, in which one or more processors are configured to execute:
>
> a parser that identifies a telematics domain from the recognized one or more words or phrases of the natural language utterance.
>
> Independent claim 71 (as amended) recites similar features.
>
> Knockeart relates to a "combined manual and spoken interface for a telematics system" that "provides a mixed graphical/manual and oral/spoken interface to a user." Knockeart, at ¶¶ 4, 38. The interface provides a "menu of a number of choices . . . includ[ing] 'navigation,' 'traffic,' 'news,' and 'emergency.'" Knockeart, at ¶ 44. A user can "select the 'navigation' choice . . . by speaking a predefined word to select the indicated choice, here the word 'select.'" Knockeart, at ¶ 44. That is, Knockeart relates to an interface that requires "limited to small numbers of words or phrases that can be more easily distinguished by a voice recognizer." Knockeart, Abstract. To select the navigation choice, the interface in Knockeart requires the user to use a "predefined word or phrase" such as "select." Knockeart, at ¶¶ 42-44. Thus, Knockeart fails to teach or suggest identifying a telematics domain from one or more words or phrases of a natural language utterance. Moreover, Arnold does appear to cure at least the foregoing deficiencies of Knockeart. Thus, Knockeart and Arnold, taken either alone or in combination, fail to teach or suggest each and every feature of independent claims 56 and 71.

For at least the foregoing reasons, withdrawal of the rejection of independent claims 56 and 71 is earnestly sought.

(Applicant's Response, p. 15.)

The patentability of the dependent Claims is argued based on their dependence from the independent claims. (Applicant's Response, p. 15.)

The above appears to argue that the NLP of claim is distinguished from limited vocabulary speech recognition of Knockeart which is based on menu of limited choices. The secondary reference, Arnold, is summarily dismissed by the above arguments. Arnold was cited for speech recognition and natural language processing of user query. (See, e.g., Arnold, Figure 21 and Col. 20, lines 5-12: "Each user query or search parameter can be entered using the keyboard or pointing device of the handheld computer 2010 or 2014. Alternatively, the user can verbally instruct the handheld computer 2010 or 2014 with the query or search parameter.") If the result of the NLP of Arnold matches a command of Knockeart, then the telematics domain is identified from the natural language utterance and the command may be carried out.

In view of the amendments to the claims, new grounds of rejection are provided.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 56, 68, 71, 83, 86, 88-89, 91-95, and 99 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold (U.S. 6,910,003) in view of Kuzunuki (U.S. 6,944,583).

Regarding **Claim 56**, **Arnold** teaches:

56. A mobile system for processing natural language utterances, comprising: [Arnold teaches a NLP processing system. Figure 20, 1100. The system of Arnold is not mobile but it serves mobile users that include speech recognition. Figure 20, 2010, 2014.]

one or more processors configured to: [Arnold, Figure 22. "Preferably, the invention is implemented in a computer program executing in a computer system. Such a computer system may include a processor, a data storage system, at least one input device, and an output device. FIG. 22 illustrates one such computer system 2200, including a processor (CPU) 2202, a RAM 2204, a ROM 2206 and an I/O controller 2210 coupled by a CPU bus 2208." Col. 22, 10-13.]

receive a natural language utterance; [Arnold, Figure 21, "speech recognizer 2102." "... wherein the user inputs the query in natural language." Col. 41, lines 1-2. "The method should allow a user to use natural language, and search based on idea concepts, rather than strict Boolean strings." Col. 2, lines 37-39 (Background). "A query is received from a user, which, as mentioned above, may be input in natural language." Col. 3, lines 18-20.]

execute a speech recognition engine that recognizes one or more words or phrases in the natural language utterance, [Arnold, Figure 21, "speech recognizer 2102." Figure 20, "Alternatively, the user can verbally instruct the handheld computer 2010 or 2014 with the query or search parameter. In this case, the handheld computer 2010 or 2014 can execute a speech recognizer 2008 that maps the user's voice to a set of reference patterns ...." Col. 20, lines 7-13.]

execute a parser that identifies a telematics domain **from the recognized one or more words or phrases of the natural language utterance and formulates a request based on the** telematics domain; and [Arnold, teaches the parsing of the NL query and also teaches formulating a database query based on the user NL query; it also teaches identifying the topic of the query. But it does not teach identifying a telematics domain. Figure 4, "natural language question parser" as part of the "user interaction module 406." Figure 15: "First, the user submits a natural language query to the system 1100 (step 1502). As discussed earlier, an exemplary natural language query can be "Tell me about joint ventures involving SBC in the Communications Services Sector" or "Did Barnes & Noble buy anyone this year?" The natural language query is parsed by the natural language user interface 1110. A variety of natural language parsers can be used. … The output from executing the query grammar files is used to *generate database queries* with one or more search parameters *corresponding to the natural language queries*." Col 17, line 57-Col. 18, Lines 3. "Once the system 1100 receives a natural language query from the user, natural language rules are applied to understand the specific goals of the question: the topic area, the type of information sought, and any specific constraints (step 1504)." Col. 18, Lines 16-20.]

communicate the request to a telematics control unit **that is configured to process the request based on the** telematics **domain**, [Arnold, Figure 11, "the query and reply front end 1120" and the "communication engine 1112." "The query and reply front end 1120 converts user queries, which can be natural language queries, or search terms, into an internal query format and submits the query to the information extraction and query engine 1104." Col. 13,

Lines 29-33. "The communication engine 1112 also receives input from a natural language query user interface 1110." Col. 13, lines 26-27.]

wherein the one or more processors are part of the telematics control unit **or are separate** from the telematics control unit. [Arnold, Figure 22 shows the processors 2202 that are part of the main system (Col. 22, row 14) and Figure 20 shows the "mobile devices 2010, 2014" that are in communication with the "natural language query system 1100" and "Each of the mobile devices or handheld computers 2010 and 2014 has a processor …." Col. 19, lines 61-62. Arnold also shows that the mobile device 2010 has speech recognizer "SR 2008." So, Arnold covers the central or distributed processing of this claim element but does not teach that the central system (1100) is a "telematics control unit."]

Arnold teaches the parser of the claim and also teaches identifying the topic/context/domain of a parsed query. Arnold does not mention the telematics domain but it can identify topic/context/domain and telematics is one topic/context/domain.

Kuzunuki separates command from query and includes a telematics domain as one type of its "Application Service Providers" and teaches:

execute a parser that **identifies a telematics domain from the recognized one or more words or phrases of the natural language utterance and formulates a request based on the telematics domain**; and [Kuzunuki, Figure 1, "mobile car PC 10c" and "navigation information ASP 60." "The speech portal server 50 recognizes the speech, applies the command conversion to the content of the speech for converting it into a command and an object to be searched for, and transmits them to an ASP 60 corresponding to the content of the command." Col. 5, lines 15-20.]

communicate the request to a telematics control unit that is configured to process the request based on the telematics domain; [Kuzunuki, Figure 8, "conversation control means 507" and Figure 17 showing the "ASP control 507d" performed by the "conversation control means 507." "The basic conversation rules 507b store rules commonly used among the individual information ASP's, and conversation rules specific to the individual information ASP's are downloaded from the individual information ASP'S. When the conversation processing mean 507a receives Vtext2 of output from the speech recognizing mean 505, determines a command ID, determines which information ASP it corresponds to, and transmits the command ID and an object as ASPDataOut to the corresponding information ASP." Col. 9, lines 1-10.]

wherein the one or more processors are part of the telematics control unit or are separate from the telematics control unit. [Kuzunuki, the APSs including the "navigation information ASP 60a" of Figure 1 and the "Mobile CarPC 10c" each have their own processor.]

Arnold and Kuzunuki pertain to the art of speech. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the telematics-related capability of Kuzunuki (which also includes natural language processing but is less elaborate in spelling out the parsing and other NLP steps) with the natural language processing of Arnold and expand the capability of Arnold to distinguishing speech that is directed at telematics.

Regarding **Claim 68**, Arnold/Kuzunuki teach the speech enabled telematics system, as applied to Claim 86, and **Kuzunuki** also teaches interactive guidance based on location:

68. The mobile system of claim 86, **wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance relating to one or more tasks for a person in the vehicle based on a location of the vehicle**. [Kuzunuki, Figure

25, command of navigation information elicits path information and map information based on input of the current position. See Col. 11, lines 13 to Col. 12, line 4.]

The combination would have been obvious according to the rationale provided for Claim 56 and because the interactive guidance feature is a feature of GPS.

Apparatus Claim 56 and method **Claim 71** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 71 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

Apparatus Claim 68 and method **Claim 83** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 83 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

Regarding **Claim 86**, **Arnold** teaches:

86. The mobile system of claim 56, wherein the one or more processors are configured to receive the natural language utterance from a speech unit operatively coupled to a computer device in a vehicle, and [Arnold receives natural language input and it may be used in a car; Figure 4, Mobile CarPC.]

wherein the telematics domain is a query and response domain or is a control domain. [Arnold teaches the query and response portion of this claim in Figure 1 and Figure 4 where navigation information is queried based on the input to the MIC 10c4 and the "position detecting means 10c8." Arnold teaches the command portion in Figure 25 by "path search command" which includes a command portion and also a query based on the command.]

Regarding **Claim 88**, **Arnold** teaches:

88. The mobile system of claim 86, wherein the parser is configured to identify the telematics **domain further from a context of the natural language utterance.** [Arnold teaches the use of domain dependence (col. 31, line 45) and context dependence (col. 32, line 22) in its speech processing and in particular in parsing (Col. 32, lines 3-9).]

Arnold does not teach the telematics domain.

Kuzunuki teaches that the domain of a speech, including the telematics (GPS, Car MobilePC) is identified from the commands that are extracted from the input. (Kuzunuki, Figure 8.)

It would have been obvious to combine the feature of Kuzunuki that looks for a telematics context/domain with the parser of Arnold that is directed to natural language processing in order to include telematics as one of the domains to be identified.

Regarding **Claim 89**, **Arnold** teaches:

89. The mobile system of claim 88, wherein the parser is configured to determine the context based on a user profile associated with the natural language utterance **or based on a previously received natural language utterance.** [Arnold teaches that context means "previous utterance." Col. 32, line 22.]

Regarding **Claim 91**, Arnold does not teach about distinguishing commands and queries. **Kuzunuki** teaches:

91. The mobile system of claim 88,

wherein the one or more processors are configured to determine, based on the context, whether the received natural language utterance is part of a query request or part

of a command request, and [Kuzunuki, teaches distinguishing command from query (Figures 8-15 and Abstract) based on the duration of the utterance or the number of characters in it. Then, based on context of the command, the type of command and its objects are identified (Figure 15).]

wherein the telematics control unit is configured to retrieve information responsive to the query request or cause an action consistent with the command request**.** [Kunuzuki, Figure 25 shows the information that is retrieved regarding the path and map responsive to the query.]

Combination is motivated according to the rationale of Claim 56.

Regarding **Claim 92**, Arnold does not teach the telematics domain but **Kuzunuki** teaches:

92. The mobile system of claim 86,

wherein the parser is configured to identify, from among a plurality of domain agents, a domain agent that is associated with the telematics domain and is configured to formulate the request based on the identified domain agent, and [Kuzunuki, Figure 8 and Figure 9.]

wherein the telematics control unit is configured to process the request by executing the identified domain agent. [Kuzunuki, Figure 9, Figure 14, Figure 15 and Figure 25 where the operation of the Mobile CarPC for navigation is shown.]

Combination is motivated according to the rationale of Claim 56.

Regarding **Claim 93**, Arnold teaches the different domains and contexts and **Kuzunuki** teaches:

93. The mobile system of claim 86, wherein the one or more processors are configured to receive a second natural language utterance from the speech unit, and wherein the parser is configured to identify a non-telematics domain from one or more words or phrases of the second natural language utterance. [Kuzunuki, Figure 1, several different ASPs (application service providers) are provided that each pertain to a different domain including non-telematics domains.]

Combination is motivated according to the rationale of Claim 56.

Regarding **Claim 94**, **Arnold** teaches:

94. The mobile system of claim 86, wherein the one or more processors are further configured to receive a non-voice input from the computer device in the vehicle, wherein the parser is configured to formulate the request further based on the non-voice input. [Arnold teaches text input which can be text but also accommodates speech that is converted to text and is then parsed. (Figure 21, multimedia data source). "FIG. 1 depicts an illustrative query input page and an answer page according to one embodiment of the present invention;" Col. 3, lines 62-64.]

Regarding **Claim 95**, Arnold does not but **Kuzunuki** teaches:

95. (New) The mobile system of claim 86,

wherein the telematics control unit is operable to interface with a vehicle-related system, the vehicle-related system comprising a vehicle entertainment device, a communication device in the vehicle, a vehicle control system, a vehicle power management system, a vehicle sensor, a control device external to and in wireless communication with the vehicle, or a vehicle history storage device. [Kuzunuki, the "Mobile CarPC" is a "vehicle-

related system" which comprises an "onboard terminal 10c" which is a communication device in the vehicle.]

Combination is motivated according to the rationale of Claim 86.

Regarding **Claim 99**, Arnold Figure 20 and Figure 21 show the speech unit but do not show that it is in a car. **Kuzunuki** teaches:

99. The method of claim 71, wherein receiving the natural language utterance comprises receiving a natural language utterance at a speech unit operatively coupled to a computer device in a vehicle. [Kuzuniki, Figure 4, Mobile CarPC and the MIC 10c4 are in a car.]

Combination is motivated according to the rationale of Claim 56.

5. Claims 57 and 72 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki in view of Sukehiro (2004/0205671).

Regarding **Claim 57**, Arnold/Kuzunuki teach a telematics control unit, as applied to Claim 56, but do not teach a multiuser situation with overlapping and interleaved dialogs. **Sukehiro** teaches managing of multiple users in the context of translation requests to translation systems instead of telematics control unit of the claim. Sukehiro also teaches that the multiple users are served simultaneously (i.e., overlapping/interleaving):

57. The mobile system of claim 56, wherein the one or more processors are configured to execute **an event manager providing a multi-threaded environment for processing a plurality of requests in a plurality of different domains across a plurality of overlapping or interleaved dialogs**, [Sukehiro, Figure 10, "communication control unit 31," "[0110] The machine translation network system 1 in this embodiment is capable of responding to translation requests from multiple users

simultaneously. A single paired machine translation system 14 and HTML parser 17 can operate on a time-sharing basis to respond to multiple translation requests simultaneously, for example, or the system may include multiple pairs of these facilities, which respond to separate translation requests simultaneously. In the latter case, multiple translation requests can be handled simultaneously by loading copies of a machine translation program into the main memories of multiple central processing units (CPUs) with which the server 3 is provided." "[0140] The communication control unit 31 (which includes functions of the HTTP daemon 10, log analyzer 11, communication tools 13a, translation processor 13b, and user registration and authentication unit 13d in FIG. 1) controls communication with the browser and input devices and an external Internet facility (not visible) that stores documents, enabling the retrieval and translation server 23 to retrieve documents from the external Internet facility and supply the retrieved documents or translations thereof to the browser and input devices 24." "[0139] The retrieval and translation server 23 retrieves documents and executes various tasks, including machine translation of the documents. Its component elements include a communication control unit 31, a machine translation unit 32, a dictionary manager 33, a dictionary data base 34, and a terminology incorporator 35."]

wherein the event manager coordinates interaction between the parser and the telematics control unit **to enable the** telematics control unit **to process the request during the overlapping or interleaved dialogs.** [Sukehiro, Figure 1 and Figure 10, "communication control unit 31" in Figure 10 which includes 10, 11, 13a and 13b in Figure 1 is shown in Figure 1 to coordinate between "dictionary manager 33", the "machine translation unit 32" and the "browser and input devices 24". The material input in Sukehiro has to be parsed before it is translated. Sukehiro is translation related instead of telematics related. So, in the combination of prior art pieces, the task done in the context of translation in Sukehiro is combined with the context of telematics from other prior art.]

Sukehiro is directed to natural language processing as are Arnold and Kuzunuki and it would have been obvious to combine the multi-user capability of Sukehiro that is used in the context of machine translation with the telematics control of Arnold/Kuzunuki to arrive at the "event manager" of the claim.

Apparatus Claim 57 and method **Claim 72** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 72 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

6. Claims 58 and 73 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki in view of Knockeart (U.S. 6,968,311).

Regarding **Claim 58**, Arnold and Kuzunuki do not teach what happens in the case of sensing hazardous conditions.

Knockeart teaches:

58. The mobile system of claim 86, **wherein the telematics control unit is configured to determine whether processing the request will create a hazardous condition for the vehicle**. [Knockeart, teaches that telematics requests that are deemed dangerous are rejected: "Device 100 accepts external condition signals 180. Depending on those signals, in some modes of operation device 100 rejects manual inputs. A "switch" 360 in the flowchart in FIG. 3 indicates this selection of modes in which manual input is rejected. For example, in some configurations in which device 100 is integrated in a vehicle (see FIG. 2), manual inputs are rejected when the vehicle speed is too fast for the user to safely use manual inputs and manual input could result in a user being distracted for too long a time at that speed. Note that whether or not manual inputs are accepted may depend on the nature of those inputs. For example, while it may be safe to provide a simple manual input for which the user does not have to take his eyes off the road such a manual input may not be rejected, but a manual input that requires the driver to select from a list or type an input on a numeric or alphabetic keyboard may be rejected since this would require the driver to potentially unsafely take his eyes off the road.." Col. 4, Lines 47-64.]

Knockeart, Arnold, and Kuzunuki are from the art speech and it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the telematics related

aspects of Knockeart that pertain to hazards of driving with the speech related commands of Arnold/Kuzunuki to arrive at a speech driven command/query system that performs a command in view of determining the potential outcome considering the condition of operation of the car.

Apparatus Claim 58 and method **Claim 73** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 73 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

7. Claims 59 and 74 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki and Knockeart in view of Lemelson (2002/0022927).

Regarding **Claim 59**, Knockeart teaches the telematics control unit that prevents certain types of manual input in response to detecting a high speed of travel (Col. 4, lines 46-64). Knockeart is not explicit about what it does in terms of communicating the hazard to the user other than cutting off his manual input option.

Lemelson teaches:

59. The mobile system of claim 58, wherein the telematics control unit is further configured to:

provide interactive guidance to resolve the hazardous condition via an output device connected to the vehicle, wherein the telematics control unit provides the interactive guidance to resolve the hazardous condition in response to determining that processing the request will create the hazardous condition; and [Lemelson, teaches that an accident situation is determined based on position of the vehicle and other data and is communicated to the driver: " GPS satellite (4) ranging signals (6) received (32) on comm1, and DGPS auxiliary range

correction signals and pseudolite carrier phase ambiguity resolution signals (8) from a fixed known earth base station (10) received (34) on comm2, at one of a plurality of vehicles/aircraft/automobiles (2) are computer processed (36) to continuously determine the one's kinematic tracking position on a pathway (14) with centimeter accuracy. That GPS-based position is **communicated** with selected other status information **to each other one of the plurality of vehicles** (2), to the one station (10), and/or to one of a plurality of control centers (16), and the one vehicle receives therefrom each of the others' status information and kinematic tracking position. Objects (22) are detected from all directions (300) by multiple supplemental mechanisms, e.g., video (54), radar/lidar (56), laser and optical scanners. Data and information are computer processed and analyzed (50, 52, 200, 452) in neural networks (132, FIGS. 6-8) in the one vehicle to identify, rank, and evaluate collision hazards/objects, an expert operating response to which is determined in a fuzzy logic associative memory (484) which generates control signals which actuate a plurality of control systems of the one vehicle in a coordinated manner to maneuver it laterally and longitudinally to avoid each collision hazard, or, for motor vehicles, when a collision is unavoidable, to minimize injury or damage therefrom. The **operator is warned** by a heads up display and other modes and **may override**. An automotive auto-pilot mode is provided." Abstract. See also paragraph 54.]

process the request in response to the computer device receiving an input that manually overrides the hazardous condition determination. [Lemelson, " … The operator is warned by a heads up display and other modes and may **override**. An automotive auto-pilot mode is provided." Abstract. See also paragraph 54.]

Lemelson like the combination Arnold/Kuzunuki/Knockeart is directed to telematics and control of vehicles. Accordingly, it would have been obvious to combine Lemelson with the combination to arrive at a device with the added safety feature of accident prevention.

Apparatus Claim 59 and method **Claim 74** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 74 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

8. Claims 60, 62, 65, 67 and 75, 77, 80, 82, and 98 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki in view of ObradovichI (7,902,969).

Regarding **Claim 60**, Kuzunuki teaches receiving GPS data from the care (Figure 4) but Arnold and Kuzunuki do not particularly teach that the sensors receive diagnostic information.

ObradovichI teaches:

60. The mobile system of claim 86, wherein the telematics control unit is configured to:

receive diagnostic information from one or more sensors connected to the vehicle, wherein the diagnostic information identifies one or more problems with the vehicle; [ObradovichI, Figure 1, Abstract: "In accordance with the invention, the system provides driving assistance to the user by taking into account the user's physical condition, the vehicle condition and the surrounding conditions.... The vehicle condition concerns the conditions of the brakes, steering, tires, radiator, etc. of the vehicle."]

communicate the one or more problems via an output device connected to the vehicle; and [ObradovichI, Figure 20, e.g., "FIG. 20 illustrates the polling routine for monitoring

the engine temperature. Instructed by such a routine, processor 103 at step 1603 retrieves temperature data from temperature subsystem 160..... If it is determined that the engine temperature is dangerously high, processor 103 at step 1614 **warns the user** about the engine overheating condition. At step 1617 processor 103 further instructs the user on how to rectify the condition. For example, instructions such as "Turn on Heater" to assist in cooling down the engine, "Open Windows" to prevent the user from getting too hot inside the car, and "Drive to a Service Station" may be communicated through display 117 and/or user interface 119." Col. 21, Lines 33-47. "User interface 119 comprises conventional audio circuitry including a microphone, voice recognition circuit, voice synthesizer and speaker to allow communications of verbal commands and audio information with system 100." Col. 4, Line 66 to Col. 5, Line 3.]

communicate with one or more systems that relate to resolving the one or more problems with the vehicle. [ObradovichI, "If the temperature data indicates that the engine is dangerously close to overheating or has overheated, processor 103 may instruct the user to "Pull Over" immediately at a safe location." Col. 21, Lines 54-57.]

ObradovichI like the combination of Arnold/Kuzunuki pertains to interaction with motor vehicles and uses speech recognition/synthesis in the process. Accordingly, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Obradovich's extensive capabilities with the combination to include the capabilities of ObradovichI in the telematics control of the combination. Note that Obradovich's "user interface 119" in Figure 1 includes speech recognition and synthesis capability. (ObradovichI, Col. 4, Line 66 to Col. 5, Line 3: "User interface 119 comprises conventional audio circuitry including a microphone, voice recognition circuit, voice synthesizer and speaker to allow communications of verbal commands and audio information with system 100.")

Regarding **Claim 62**, Arnold/Kuzunuki, as applied to Claim 86, teach the speech controlled telematics control system but do not teach the specifics of this claim that **ObradovichI** teaches:

62. The mobile system of claim 86, wherein the telematics control unit is configured to:

detect an accident situation from information received from one or more sensors connected to the vehicle; [ObradovichI, Figure 23, "FIG. 23 illustrates a safety routine for determining whether the vehicle would clear an object in its path. Such a safety routine is triggered when one such object is detected, e.g., by detection subsystem 130." Col. 23, Lines 48-51. If it does not clear the object, it will collide with it.]

provide one or more interactive announcements or warnings via an output device connected to the vehicle in response to detecting the accident situation, receive an input that is in response to the one or more interactive announcements or warnings, the input indicating a nature of the accident situation or a condition of a person in the vehicle; and [ObradovichI, Figure 23 "Processor 103 at step 1915 determines whether the lowest vehicle clearance exceeds the height of the object. If so, the routine comes to an end. Otherwise, processor 103 at step 1918 generates a **warning** to be shown on display 117 and/or **announced** through user interface 119, and instructs the user to either negotiate around the object or stop." Col. 23, Lines 57-65.]

communicate the nature of the accident situation or the condition of the person in the vehicle to emergency personnel over a communications link. [ObradovichI, The scenario presented in Figure 23 does not expressly include a communication to emergency personnel. However, according to the description of Figure 4, the impaired condition of the driver is

communicated to the emergency personnel: "Referring to FIG. 4, recorder subsystem 178 is provided for recording data related to the driver's condition and behavior, the vehicle's condition, and the surrounding conditions..... Processor 2303 may communicate the data to a remote location, via communications subsystem 154. The remote location receives the data from which information is derived for use in reports for fleet management, **police**, insurance, and/or other purposes." Col. 6, Lines 41-59. Police constitutes emergency personnel.]

Same motivation to combine as provided for Claim 60.

Regarding **Claim 65**, Arnold/Kunuzuki, as applied to Claim 86, teach the speech controlled telematics control system but do not teach the specifics of this claim that ObradovichI teaches:

65. The mobile system of claim 86, **wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance for performing one or more maneuvers to operate the vehicle via an output device connected to the vehicle**. [ObradovichI, tells the driver how to maneuver around an object on the road: Figure 23, "Processor 103 at step 1915 determines whether the lowest vehicle clearance exceeds the height of the object. If so, the routine comes to an end. Otherwise, processor 103 at step 1918 generates a warning to be shown on display 117 and/or announced through user interface 119, and instructs the user to either negotiate around the object or stop." Col. 23, Lines 57-65.]

Same motivation to combine as provided for Claim 60.

Regarding **Claim 67**, Arnold/Kunuzuki, as applied to Claim 86, teach the speech controlled telematics control system but do not teach the specifics of this claim that ObradovichI teaches:

67. The mobile system of claim 86, wherein the telematics control unit is further configured to:

detect that the vehicle is approaching or has exceeded a travel limit for the vehicle; [ObradovichI, Figure 5 includes a situation where the location of the driver at time t may be determined. From this information it can be detected whether the driver has exceeded a travel limit: "It should be realized that additional data can be analyzed when determining the context of the driver's situation at time t. For example, navigation data from vehicle control subsystem 121 can be used to determine the vehicle's location at time t and indicate whether the driver was on course at that time." Col. 7, Lines 21-26.]

provide interactive guidance for returning to the travel limit via an output device connected to the vehicle in response to detecting that the vehicle is approaching or has exceeded the travel limit; and [ObradovichI, "When the user utilizes navigation system 205 to request instructions for a given destination, the user is elicited for information concerning the destination, any intermediate stops, etc. Such elicitation is realized by posing questions on display 117 and/or by uttering those questions using a synthesized voice through the speaker in user interface 119. The user then provides verbal responses (or commands) to such questions through the microphone in same. Relying on the aforementioned speech recognition circuitry, navigation system 205 recognizes and registers the responses. Using stored map information, system 205 then provides on display 117 a suggested route leading to the destination. Furthermore, based on the knowledge of the vehicle's instantaneous speeds and directions, system 205 is capable of verbally and visually directing the user to the destination." Col. 5,

Lines 39-53. Part of directing the user to the destination is telling the user where he has arrived or has past the destination.]

determine whether to extend the travel limit. [ObradovichI, Col. 5, Lines 39-53, provided above. The user can provide verbal commands to the navigation system. The user can provide a new destination which would be equivalent to extending the travel limit.]

Same motivation to combine as provided for Claim 60.

Apparatus Claims 60, 62, 65, and 67 and method **Claims 75, 77, 80, and 82** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claims 75-77, 80 and 82 are similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

Regarding **Claim 98**, Arnold/Kunuzuki, as applied to Claim 86, teach the speech controlled telematics control system but do not teach the fatigue aspects that **ObradovichI** teaches:

98. (New) The mobile system of claim 86**, wherein the system is configured to receive an indication of fatigue of a user of the vehicle and to initiate a dialog to determine the user's fatigue level**. [ObradovichI "Thus, in accordance with an aspect of the invention, system 100 includes driver condition subsystem 123 to provide information concerning the physical condition and cognitive state of the user handling the vehicle. Referring to FIG. 3, subsystem 123 comprises cognitive measuring system 303 for measuring the user's alertness, which may be adversely affected by illness, extreme **fatigue**, abnormal stress, medication or alcohol. For example, to determine whether the user is alert enough to drive, when the user attempts to start

the subject vehicle, system 303 causes different colors in random order to be shown one by one on display 117. The user is requested to identify each displayed color via voice recognition or selection of an appropriate choice shown on display 117. If each displayed color is correctly identified, processor 103 causes the vehicle to start."]

Same motivation to combine as provided for Claim 60.

9. Claims 96 is are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki in view of ObradovichI and Breed (US 2002/0188392).

Regarding Claim 96, Arnold/Kuzunuki/ObradovichI, as applied to Claim 62, teach detecting of an accident and **Breed** teaches:

96. (New) The mobile system of claim 62, **wherein the telematics control unit is configured to communicate medical information associated with the person in the vehicle over the communications link or via an output device in the vehicle**. [Breed, " [0018] The present invention is related to methods and apparatus for sensing and determining the status of occupants in a vehicle prior to, during and/or after an accident or deployment of an airbag and transmitting data relating to the occupants to a remote facility such as an emergency response station."]

Breed pertains to telematics as does the combination of Arnold/Kuzunuki/ObradovichI and it would have been obvious to add the reporting of the situation of the passengers to the medical staff if an accident is sensed considering that the passengers might be incapacitated.

10. Claims 63 and 78 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold, Kuzunuki and ObradovichI, in view of Odinak (7,406,421).

Regarding **Claim 63**, ObradovichI, as applied to Claim 62, teaches the detection of an emergency situation and establishing communication with the police but is not explicit regarding a voice channel. **Odinak** teaches this feature:

63. The mobile system of claim 62, **wherein to the telematics control unit is configured to establish a voice channel over a communications link to enable communication between the person in the vehicle and the emergency personnel**. [Odinak, Figure 1 shows the components and Figure 5 shows that the user can contact a human operator if he wants to (element 220: "Does the user want to contact a human operator?" and "connect user to the human operator 224." "FIG. 1 shows a system 20, a telematics control unit, within a vehicle that performs the present invention. System 20 includes a user interface 22 having a microphone 24 for capturing the user's voice; a display 26; speakers 28; and an interface entry 30 for allowing the user to perform various interactive functions. System 20 also includes a processor 32; a global positioning system (GPS) 34 for determining precise vehicle locations; and a communications device 36, such as a cellular modem, for transmitting and receiving wireless information. System 20 is also coupled to receive information from other vehicle systems (not shown), such as a speedometer, an engine, or a braking system."]

Odinak is from the art of speech applied in "a vehicle based telematics control unit" as Arnold/Kuzunuki/Obradovich combination is. It would have been obvious to one of ordinary skill at the time of the invention to combine the capability of Odinak regarding live communication with police and emergency personnel in the case of an emergency where time is of the essence.

Apparatus **Claim 78** and method Claim 63 are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly Claim 78 is similarly rejected under the same rationale as applied above with respect to apparatus Claim 63.

11. Claims 64 and 79 are rejected under 35 U.S.C. 103(a) as being unpatentable over Knockeart and Arnold, in view of ObradovichII (6,275,231).

Regarding **Claim 64**, Arnold and Kuzunuki teach voice recognition in the context of voice controlled devices in telematics but do not address panic and emergency situations. ObradovichII teaches a manual panic/emergency key in a car:

64. The mobile system of claim 56, **wherein to process the request in the telematics control unit, the telematics control unit communicates over a communication link an indication that the one or more** recognized **words or phrases include at least one predetermined panic word, panic phrase, emergency word, or emergency phrase**. [ObradovichII, Figure 2 has different keys that may activate different operations: "When emergency key 211g is depressed, processor 103 causes emergency options to be displayed on display 205. These options may include police-type emergencies, medical emergencies, mechanical problems, panic measures, etc. For example, upon selection of the police-type emergencies option, the user may be queried whether he/she is a victim. If the user responds affirmatively, processor 103 immediately establishes a phone or modem connection with a police authority to automatically furnish such preliminary information as the identity of the user, description of the vehicle and emergency contacts, which are pre-recorded, along with the current location of the vehicle identified by navigation system 329. If possible, the user may also

add, in the communication, the nature of the emergency and current statuses of the user and any passengers. By automatically communicating at least the preliminary information to the proper authority in case of an emergency, the user can be instantly reached and helped." Col. 10, line 66 to Col. 11, Line 16. See also Col. 11, Lines 17-23 for a panic alarm.]

ObradovichII like the combination of Knockeart/Arnold pertains to control of motor vehicles and uses speech recognition/synthesis in the process. Accordingly, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine ObradovichII's panic and emergency features with Knockart/Arnold to address such situations that are an unfortunate aspect of driving.

Apparatus Claim 64 and method **Claim 79** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly Claim 79 is similarly rejected under the same rationale as applied above with respect to apparatus Claim 64.

12. Claims 66 and 81 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold, Kuzunuki and ObradovichI in view of Huang (6,362,748).

Regarding **Claim 66**, Arnold/Kuzunuki teach the speech controlled telematics system of Claim 86 and **ObradovichI** teaches:

66. The mobile system of claim 86, wherein to process the request in the telematics control unit, the telematics control unit:

provides interactive routing information to coordinate a rendezvous with one or more other vehicles in communication with the vehicle; and [ObradovichI teaches a two-way communication with other vehicles regarding road conditions ("It should be noted that the

visibility measurement in the subject vehicle may be broadcast to other nearby vehicles using communications subsystem 154 (described below) so that they can benefit from such a measurement. Indeed, the user of the subject vehicle may be a beneficiary of such visibility information received from another vehicle traveling in the same local area." Col. 14, lines 42-52. See also Col. 2, Lines 6-17.) This kind of information includes location information of the vehicle that is reporting the data because it has to tell you where on the road the visibility is going to be bad. ObradovichI thus implicitly teaches reporting the location of a car to other cars.]

establishes a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles.

The combination of ObradovichI and Arnold/Kuzunuki is obvious under the rationale of Claim 60.

Arnold/Kuzunuki and ObradovichI do not expressly teach calling other cars.

Huang teaches this feature:

establishes a voice channel over a communications link to enable communication between a person in the vehicle and one or more people in the other vehicles. [Huang, "The license plate number of a specific vehicle is related to a passcode that is built into a communications reception unit installed in the vehicle. When a driver enters a plate number of the vehicle having the communication reception unit installed, he can communicate with another vehicle. The communication reception unit can receive satellite positioning signals to provide more accurate coordinates to a control center, which can determine the position of the vehicle and send accurate, useful and relevant information to the vehicle." Abstract. "It is thus a

primary objective of the invention to provide a system for communicating among vehicles that enables moving vehicles to communicate with each other and can send accurate coordinates to a control center." Col. 1, Lines 56-60. See also Figure 1 and Col. 4, Lines 50-65.]

Huang pertains to communications from a car and during driving and is thus related to telematics of Arnold/Kuzunuki/ObradovichI combinations. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the cell phone calling capability of Huang with the combination of other references to reduce the danger of diversion of attention during driving.

Apparatus Claim 66 and method **Claim 81** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 81 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

13. Claims 69 and 84 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki, in view of Wolfe (6,877,001).

Regarding **Claim 69**, Arnold/Kuzunuki teach the speech enabled telematics system, as applied to Claim 56, but do not teach the speech recognizer with updatable dictionary. **Wolfe** teaches "retrieving documents with spoken queries" and keeps its speech recognition dictionary by updating it using the keywords from previously retrieved results (documents) which correspond to previous spoken queries. This way it has a vocabulary that is not static and yet it is not too large:

69. The mobile system of claim 56**, wherein the speech recognition engine recognizes the words or phrases in the natural language utterance using a plurality of dictionary and**

phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs. [Wolfe, Figure 1, "Detect keywords and key phrases and construct vocabulary 200," "Dynamic vocabulary and grammar 151" back to the "speech recognition engine 150." "FIG. 2 is a flow diagram of a method for constructing a dynamic speech recognizer vocabulary for an information retrieval system ...." Col. 3, Lines 17-19. "In addition, an active vocabulary and grammar of the speech recognizer or search engine are dynamically updated with key words and key phrases that are automatically extracted from the documents as they are indexed. In other words, information from the document index is fed back into the recognizer or search engine itself." Col. 3, Lines 1-12.]

Wolfe is related to the art of speech as is the combination of Arnold/Kuzunuki and it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the updatable dictionary feature of Wolfe with the combination of Arnold/Kuzunuki to provide a dictionary that is not too massive yet applicable. (See, e.g., Wolfe, paragraph 10).

Apparatus Claim 69 and method **Claim 84** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claim 84 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

14. Claims 70 and 85 are rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold, Kuzunuki in view of Wolfe and Shen (7,478,036).

Regarding **Claim 70**, Wolfe teaches that the probabilities of the query words are taken into account in the certainty of the recognition results which in turn generates the search results

which yield the dynamic dictionary. [Wolfe, "The invention provides a system and method for retrieving documents from a multi-media database using spoken queries. In addition, the invention makes use of document index information in the speech recognition process, and certainty information about the recognition result while searching for matching documents in the database. The certainty information represents the probabilities of possible query words." Col. 3, Lines 24-59.] Wolfe is not as explicit as **Shen** regarding the use of apriori probabilities that is set forth in Shen in a more straight forward manner:

70. The mobile system of claim 69, **wherein the plurality of dictionary and phrase entries are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs**. [Shen "C. Filtering out rules may be established based upon prior probabilities. For example, if there is a prior statistic language model derived from standard corpus where we can get Ps(w1 . . . wn), which is the probability of the new extracted word NW=w1 . . . wn, we can calculate Pc(w1 . . . wn) from the current corpus easily. If Pc(w1 . . . wn)/Ps(w1 . . . wn) is large, it means this new word appears relatively more frequently in the current corpus than in standard corpus. So it is a real new word for this domain. Otherwise, it means this new word combinations are common for standard domain also, so not a new word." Col. 6, Lines 25-35.]

Shen is from the same field of speech as Arnold/Kuzunuki/Wolfe and it would have been obvious to use the prior probability of the terms in the building of the dictionary considering that this is a usual method in the art and would be use of a known technique to improve similar devices, methods, or products in the same way.

Apparatus Claim 70 and method **Claim 85** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step.

Accordingly the method Claim 85 is similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

15. Claim 90 is rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki and in view of Kanevsky (6,236,968).

Regarding **Claim 90**, Arnold/Kuzunuki teach the speech enabled telematics system, as applied to Claim 86, but do not teach user profile which is taught by **Kanevsky**:

90. The mobile system of claim 89, wherein the user profile includes information related to a user's vehicle type, vehicle location, special vehicle equipment, commonly visited destination, address, place of business, interest, or formatting or presentation preference. [Kanevesky, "In order to produce a pertinent response, the conversational planner 107 stores information about the driver, such as the driver's tastes, hobbies, age, family, etc. in a driver profile module 121." Col. 4, Lines 9-14. "taste" is or includes "interest."]

Kanvesky like Arnold/Kuzunuki pertains to telematics and speech and it would have been obvious to combine the user profile feature of Kanevsky with the combination of Arnold/Kuzunki in order to fine tune the results to the user profile.

16. Claim 97 is rejected under 35 U.S.C. 103(a) as being unpatentable over Arnold and Kuzunuki and in view of Talmor (5,913,196).

Regarding **Claim 97**, Arnold/Kuzunuki teach the speech enabled telematics system, as applied to Claim 86, but do not teach the authentication which is taught by **Talmor**:

97. The mobile system of claim 86, wherein the system is configured to authenticate a user based on the received natural language utterance and grant vehicle access to the user in response to the authentication. [Talmor, abstract, teaches voice authentication and Col. 1, line 17 teaches that "cars" are one type of "secured-system" that use the voice authentication of Talmor.]

Talmor pertains to the art of speech as does Arnold/Kuzunuki and it would have been obvious to combine the voice authentication of Talmor with the system/method of Arnold/Kuzunuki to provide secure access to control of the car telematics.

Double Patenting

17. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory obviousness-type double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re*

Vogel, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent either is shown to be commonly owned with this application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement.

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

18. **Claims 61 and 76 are rejected on the ground of nonstatutory obviousness-type double patenting as being unpatentable over claim 48 of U.S. Patent No. 7,693,720 to Kennewick in view of Arnold (U.S. 6,910,003), Kuzunuki (U.S. 6,944,583), ObradovichI (7,902,969), Funk (2003/065427), and Breed (2002/0188392).**

Although the conflicting claims are not identical, they are not patentably distinct from each other because the element of Claims 61 and 76 that is not taught by prior art references, is claimed in claim 48 of the issued patent.

Regarding **Claim 61**, Arnold and Kuzunuki and ObradovichI, as applied to Claim 60 (shown above in the obviousness rejection of the claims), teach a mobile system that receives diagnostic information and the additional references further teach:

61. The mobile system of claim 60, wherein the one or more systems are related to scheduling service for the vehicle, [**Funk**, "[0034] Whatever system analyzes Data 36 may respond via Voice Portal 27 to Mobile Unit 28 by Information 37 which may be an instruction to adjust any number of electronically adjustable engine settings (e.g. fuel/air mixture, etc.). Alternatively, another Data Request 35 may be sent to Mobile Unit 28 to obtain more information. Result 32 sent from the analytical system may indicate that a professional service is necessary. Result 32 may also indicate a priority consisting of either immediate service required, service this week, or service this month. Additionally, Voice Portal 27 may choose a service center based on the priority of the service and a proximity to the current location if the service is immediately necessary, or may start an inquiry with User 29 concerning scheduling a service later in the week. After Dialogue 31 obtains a target service time, Voice Portal 27 may again access the Internet or other network to schedule the service. After **scheduling the service**, Voice Portal 27 may communicate the scheduled time to Mobile Unit 28 in the form of Result 32 over Data Channel 34, which may in turn inform User 27 as well as set a reminder to remind User 27 on the day of the schedule service and/or within a prescribed time before the service."]

summoning help to the vehicle over one or more communication links, [**Breed**, " [0018] The present invention is related to methods and apparatus for sensing and determining the status of occupants in a vehicle prior to, during and/or after an accident or deployment of an airbag and transmitting data relating to the occupants to a remote facility such as an emergency response station."]

providing remedial instructions that address the one or more problems, [**ObradovichI**, teaches **remedial instructions**: "If the temperature data indicates that the engine

is dangerously close to overheating or has overheated, processor 103 may instruct the user to "Pull Over" immediately at a safe location." Col. 21, Lines 54-57. "At step 1617 processor 103 further instructs the user on how to rectify the condition. For example, instructions such as "Turn on Heater" to assist in cooling down the engine, "Open Windows" to prevent the user from getting too hot inside the car, and "Drive to a Service Station" may be communicated through display 117 and/or user interface 119." Col. 21, Lines 41-47. "If the alternator is rotating outside the predetermined range, for example, too slowly, then the user is notified that the alternator belt is slipping and should be changed or repaired." Col. 22, lines 47-50.]

ordering one or more parts for the vehicle, or providing an estimated cost for the scheduling the service for the vehicle. [**Kennewick**, U.S. 7,693,720, Claim 48: "The method according to claim 47, wherein the solution to the one or more problems includes at least one of scheduling service for the vehicle, summoning help to the vehicle over one or more communication links, providing remedial instructions that address one or more of the problems, **ordering one or more parts for the vehicle, or providing a cost estimate for the scheduled service**."]

Same motivation to combine as provided for Claim 60.

Apparatus Claim 61 and method **Claim 76** are related as apparatus and the method of using same, with each claimed element's function corresponding to the claimed method step. Accordingly the method Claims are similarly rejected under the same rationale as applied above with respect to the corresponding apparatus claim.

19. **Claim 87 is rejected on the ground of nonstatutory obviousness-type double patenting as being unpatentable over claim 5 of U.S. Patent No. 7,693,720 to Kennewick in view of Arnold (U.S. 6,910,003) and Kuzunuki (U.S. 6,944,583).**

Although the conflicting claims are not identical, they are not patentably distinct from each other because the element of Claim 87 that is not taught by prior art references, is claimed in claim 5 of the issued patent.

Regarding **Claim 87**, Arnold and Kuzunuki and ObradovichI, as applied to Claim 86 (shown above in the obviousness rejection of the claims), teach a mobile system that receives command and query and **claim 5 of Kennewick** further teaches:

87. (New) The mobile system of claim 86,

wherein the query and response domain includes a driving directions domain, a travel services domain, an entertainment scheduling domain, or a location-based marketing domain, and [Kennewick, U.S. 7,693,720, Claim 5: "5. The mobile system according to claim 1, wherein the selected domain agent includes data associated with at least one of driving directions, travel information, restaurant information, systems information for the vehicle, safety information, or entertainment information."]

wherein the control domain includes a vehicle lighting control domain, a windshield wiper control domain, or a vehicle entertainment domain. [Kennewick, U.S. 7,693,720, Claim 5.]

Conclusion

20. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Matsuda (6,108,532) prevents a call from coming to a moving car when the speed of the car is high.

ObradovichIII (6,754,485) also teaches the limitation of Claim 68:

Regarding Claim 68, Arnold/Kuzunuki teach the speech enabled telematics system, as applied to Claim 86, but do not teach interactive guidance based on location. ObradovichIII teaches an alternative to the OnStar service that uses a car email (C-mail) instead of a cell phone that is used by OnStar and provides interactive information to the driver based on GPS data of the car:

68. The mobile system of claim 86, wherein to process the request in the telematics control unit, the telematics control unit provides interactive guidance relating to one or more tasks for a person in the vehicle based on a location of the vehicle. [ObradovichIII, regarding OnStar, the background states: "The service representative may then provide the user with directions, restaurant recommendations, and/or roadside assistance depending on the vehicle's location." Col. 1, Lines 64-67. Then, the reference teaches that it would use an alternative method of communication: "In accordance with the invention, the communications between a vehicle and remote stations or servers are realized using electronic car-mail, hereinafter referred to as "C-mail."" Col. 4, Lines 33-36. Figure 1 shows the various sensors that can report on the car, Figure 2 shows the communication between the car and a central server, Figure 4 shows a "smart navigator" that can communicate with the user by touch or through a speech interface: "As shown in FIG. 4, the SMART NAVIGATOR page includes displayed options, e.g., NAVIGATION option 419, WEATHER option 421, TRAFFIC option 425 and DIRECTORY option 430. By selecting option 419 by touching same on display 405,

followed by selecting ENTER key 435, the user is able to obtain navigation instructions to a given destination. Specifically, the user is elicited for information concerning the destination, any intermediate stops, etc. Such elicitation is realized by posing questions on display 405 and/or by uttering those questions using a synthesized voice through the aforementioned audio output. The user then provides **verbal responses** thereto through the aforementioned audio input. The **speech recognizer** in interface 139 then recognizes and registers the responses." Col. 12, Lines 4-17. The "user interface 139" of Figure 1 has speech capability: "In this instance, interface 139 also includes (1) a conventional **speech recognizer** and audio input for the user to utter commands and responses to processor 103, and (2) a conventional **speech synthesizer** and audio output for processor 103 to communicate to the user in audio." Col. 11, Lines 10-13. The information provided to the driver is based on the location of the car: "Another message may contain just-in-time map information for navigation, depending on the current location of the automobile. Yet another message may contain advertising information concerning selected entities, e.g., restaurants, gas stations, department stores, etc., also depending on the current location of the automobile. The messages transmitted from the automobile to the remote server may contain, e.g., GPS information identifying the location of the automobile, dynamic data furnished by sensors in the automobile for analysis, an odometer reading and a speedometer reading for records, etc." Abstract.]

Like the combination of Arnold/Kuzunuki, ObradovichIII is directed to telematics with the use of speech. Thus, it would have been obvious to one of ordinary skill in the art at the time of invention to combined the extended capabilities of ObradovichIII regarding providing guidance to the driver with the simpler telematics system of Arnold/Kuzunuki to increase the capabilities of this system.

Pryor (2009/0273563), "[0148] … o. Preventive maintenance screen, with data, predicted service times, and even camera images of key parts of vehicle such as wheels, engine, etc. …" teaches providing the service schedule to the driver on a screen but not the actual scheduling.

Coffman (U.S. 2009/0313026) teaches the use of context in determining to which device a command is directed: " [0174] When a user interacts with the collection of devices, the interaction may always be via a central unit such as a PVA (personal vehicle assistant) 710 in a car, or a speech browser 713. The task dispatcher and context stack accumulates the contexts associated to all the devices and will parse and dispatch commands to each device accordingly. If the user interacts with the entire collection of devices, then a device is always active (the last activated context). This devices check if a new command fits its context stack. If not, it passes to a neighboring device that becomes active. The process is iterated until a match is found, and possibly confirmed by the user, or the request bounces back to the first device. In that case, an error or confusion message is returned to the user." And also "The conversational arguments allow a dialog engine to estimate the relevance of a new query by the NLU engine, based on the current state and context."

Regarding Claim 87, see **DeLine** (2009/0106029):

> a windshield wiper control domain, [DeLine, U.S. 7,693,720, "[0039] … Microphone 18 may also function as a receiver for one or more other vehicular functions and controls, such as voice activated headlamps, alarm systems, radios, cruise control, windows, cellular phones, message recorders, pagers, back up aids, windshield wipers, rain sensors and the like."] or

21. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

22. Any inquiry concerning this communication or earlier communications from the examiner should be directed to FARIBA SIRJANI whose telephone number is (571)270-1499. The examiner can normally be reached on Monday through Thursday 9am to 4pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Pierre-Louis Desir can be reached on 571-272-7799. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FARIBA SIRJANI/
Examiner, Art Unit 2626
/Michael N Opsasnick/
Primary Examiner, Art Unit 2626

Notice of References Cited	Application/Control No. 12/705,311	Applicant(s)/Patent Under Reexamination KENNEWICK ET AL.	
	Examiner FARIBA SIRJANI	Art Unit 2626	Page 1 of 4

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-5,577,165 A	11-1996	Takebayashi et al.	704/275
*	B	US-5,895,464 A	04-1999	Bhandari et al.	704/9
*	C	US-6,236,968 B1	05-2001	Kanevsky et al.	704/275
*	D	US-6,301,560 B1	10-2001	Masters, Steven P.	704/251
*	E	US-6,393,403 B1	05-2002	Majaniemi, Markku	704/275
*	F	US-2003/0065427 A1	04-2003	Funk et al.	701/1
*	G	US-6,578,022 B1	06-2003	Foulger et al.	706/45
*	H	US-2003/0110037 A1	06-2003	Walker et al.	704/257
*	I	US-2003/0120493 A1	06-2003	Gupta, Sunil K.	704/270.1
*	J	US-6,601,026 B2	07-2003	Appelt et al.	704/9
*	K	US-2003/0144846 A1	07-2003	Denenberg et al.	704/277
*	L	US-2003/0158731 A1	08-2003	Falcon et al.	704/231
*	M	US-2003/0212562 A1	11-2003	Patel et al.	704/275

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Notice of References Cited	Application/Control No. 12/705,311	Applicant(s)/Patent Under Reexamination KENNEWICK ET AL.	
	Examiner FARIBA SIRJANI	Art Unit 2626	Page 2 of 4

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-6,745,161 B1	06-2004	Arnold et al.	704/7
*	B	US-6,959,276 B2	10-2005	Droppo et al.	704/226
*	C	US-7,003,463 B1	02-2006	Maes et al.	704/270.1
*	D	US-7,069,220 B2	06-2006	Coffman et al.	704/275
*	E	US-7,107,218 B1	09-2006	Preston, Keith R	704/270
*	F	US-7,184,957 B2	02-2007	Brookes et al.	704/246
*	G	US-7,190,770 B2	03-2007	Ando et al.	379/88.01
*	H	US-7,207,011 B2	04-2007	Mulvey et al.	715/812
*	I	US-2007/0198267 A1	08-2007	Jones et al.	704/257
*	J	US-7,283,951 B2	10-2007	Marchisio et al.	704/9
*	K	US-7,299,186 B2	11-2007	Kuzunuki et al.	704/270.1
*	L	US-7,305,381 B1	12-2007	Poppink et al.	1/1
*	M	US-7,321,850 B2	01-2008	Wakita, Yumi	704/10

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Notice of References Cited	Application/Control No. 12/705,311	Applicant(s)/Patent Under Reexamination KENNEWICK ET AL.	
	Examiner FARIBA SIRJANI	Art Unit 2626	Page 3 of 4

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-7,337,116 B2	02-2008	Charlesworth et al.	704/254
*	B	US-7,328,155 B2	02-2008	Endo et al.	704/251
*	C	US-2008/0147410 A1	06-2008	Odinak, Gilad	704/270.1
*	D	US-7,406,421 B2	07-2008	Odinak et al.	704/275
*	E	US-7,447,635 B1	11-2008	Konopka et al.	704/275
*	F	US-2009/0052635 A1	02-2009	Jones et al.	379/88.02
*	G	US-2009/0106029 A1	04-2009	DeLine et al.	704/275
*	H	US-7,542,894 B2	06-2009	Murata, Masaki	704/9
*	I	US-7,558,730 B2	07-2009	Davis et al.	704/235
*	J	US-2009/0273563 A1	11-2009	Pryor, Timothy R.	345/157
*	K	US-2009/0313026 A1	12-2009	Coffman et al.	704/275
*	L	US-2010/0029261 A1	02-2010	Mikkelsen et al.	455/419
*	M	US-7,676,365 B2	03-2010	Hwang et al.	704/240

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Notice of References Cited	Application/Control No. 12/705,311	Applicant(s)/Patent Under Reexamination KENNEWICK ET AL.	
	Examiner FARIBA SIRJANI	Art Unit 2626	Page 4 of 4

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-7,693,720 B2	04-2010	Kennewick et al.	704/275
*	B	US-7,748,021 B2	06-2010	Obradovich, Michael L.	725/105
*	C	US-7,801,731 B2	09-2010	Odinak et al.	704/275
*	D	US-7,831,426 B2	11-2010	Bennett, Ian M.	704/252
*	E	US-2010/0312566 A1	12-2010	Odinak et al.	704/275
*	F	US-7,873,519 B2	01-2011	Bennett, Ian	704/257
*	G	US-7,953,732 B2	05-2011	Frank et al.	707/724
*	H	US-8,155,962 B2	04-2012	Kennewick et al.	704/257
*	I	US-8,170,867 B2	05-2012	Germain, Nicolas	704/9
*	J	US-8,219,599 B2	07-2012	Tunstall-Pedoe, William	707/603
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	1A	5,027,406	06/25/1991	Roberts *et al.*	381	43	
	1B	5,208,748	05/04/1993	Flores *et al.*	364	419	
	1C	5,357,596	10/18/1994	Takebayashi *et al.*	395	2.84	
	1D	5,577,165	11/19/1996	Takebayashi et al.	395	2.84	
	1E	5,608,635	03/04/1997	Tamai	364	449.3	
	1F	5,740,256	04/14/1998	Castello Da Costa *et al.*	361	94.7	
	1G	5,748,974	05/05/1998	Johnson	395	759	
	1H	5,839,107	11/17/1998	Gupta *et al.*	704	270	
	1I	5,897,613	04/27/1999	Chan	704	210	
	1J	5,960,397	09/28/1999	Rahim	704	244	
	1K	5,960,399	09/28/1999	Barclay *et al.*	704	270	
	1L	5,960,447	09/28/1999	Holt *et al.*	707	500	
	1M	5,995,928	11/30/1999	Nguyen *et al.*	704	251	
	1N	6,018,708	01/25/2000	Dahan *et al.*	704	244	
	1O	6,035,267	03/07/2000	Watanabe *et al.*	704	1	
	1P	6,049,607	04/11/2000	Marash *et al.*	379	410	
	1Q	6,067,513	05/23/2000	Ishimitsu	704	233	
	1R	6,085,186	07/04/2000	Christianson *et al.*	707	3	
	1S	6,108,631	08/22/2000	Ruhl	704	270	
	1T	6,167,377	12/26/2000	Gillick *et al.*	704	240	
	1U	6,173,266	01/09/2001	Marx *et al.*	704	270	
	1V	6,188,982	02/13/2001	Chiang	704	256	
	1W	6,208,964	03/27/2001	Sabourin	704	244	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO
	1X	1 320 043 A2	06/18/2003	Europe				
	1Y	1 646 037	04/12/2006	Europe				

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

	1Z	"Statement in Accordance with the Notice from the European Patent Office dated 1 October 2007 Concerning Business Methods (OJ EPO 11/2007, 592-593), XP002456252

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	07/06/2012

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

403472863v1

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	2A	6,233,561	05/15/2001	Junqua *et al.*	704	277	
	2B	6,266,636	07/24/2001	Kosaka *et al.*	704	244	
	2C	6,269,336	07/31/2001	Ladd *et al.*	704	270	
	2D	6,278,968	08/21/2001	Franz *et al.*	704	3	
	2E	6,366,882	04/02/2002	Bijl *et al.*	704	235	
	2F	6,377,913	04/23/2002	Coffman *et al.*	704	8	
	2G	6,397,181	05/28/2002	Li *et al.*	704	256	
	2H	6,405,170	06/11/2002	Phillips *et al.*	704	270	
	2I	6,429,813	08/06/2002	Feigen	342	357.13	
	2J	6,510,417	01/21/2003	Woods *et al.*	704	275	
	2K	6,522,746	02/18/2003	Marchok *et al.*	379	406.03	
	2L	6,549,629	04/15/2003	Finn *et al.*	381	92	
	2M	6,587,858	07/01/2003	Strazza	707	102	
	2N	6,594,367	07/15/2003	Marash *et al.*	381	92	
	2O	6,606,598	08/12/2003	Holthouse *et al.*	704	275	
	2P	6,622,119	09/16/2003	Ramaswamy *et al.*	704	9	
	2Q	6,633,846	10/14/2003	Bennett *et al.*	704	257	
	2R	6,681,206	01/20/2004	Gorin *et al.*	704	243	
	2S	6,701,294	03/02/2004	Ball *et al.*	704	257	
	2T	6,704,708	03/09/2004	Pickering	704	235	
	2U	6,751,612	06/15/2004	Schuetze *et al.*	707	4	
	2V	6,757,544	06/29/2004	Rangarajan *et al.*	455	456.1	
	2W	6,832,230	12/14/2004	Zilliacus *et al.*	707	203	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	07/06/2012

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	3A	6,868,380	03/15/2005	Kroeker	704	240	
	3B	6,912,498	06/28/2005	Stevens *et al.*	704	235	
	3C	6,934,756	08/23/2005	Maes	709	227	
	3D	6,954,755	10/11/2005	Reisman	707	10	
	3E	6,959,276	10/25/2005	Droppo *et al.*	704	226	
	3F	6,975,993	12/13/2005	Keiller	704	275	
	3G	6,983,055	01/03/2006	Luo	381	313	
	3H	7,003,463	02/21/2006	Maes *et al.*	704	270.1	
	3I	7,024,364	04/04/2006	Guerra *et al.*	704	270	
	3J	7,035,415	04/25/2006	Belt *et al.*	381	92	
	3K	7,043,425	05/09/2006	Pao	704	211	
	3L	7,054,817	05/30/2006	Shao	704	270	
	3M	7,058,890	06/06/2006	George *et al.*	715	728	
	3N	7,069,220	06/27/2006	Coffman *et al.*	704	275	
	3O	7,072,834	07/04/2006	Zhou	704	244	
	3P	7,107,210	09/12/2006	Deng *et al.*	704	226	
	3Q	7,110,951	09/19/2006	Lemelson *et al.*	704	270	
	3R	7,143,037	11/28/2006	Chestnut	704	251	
	3S	7,165,028	01/16/2007	Gong	704	233	
	3T	7,184,957	02/27/2007	Brookes *et al.*	704	246	
	3U	7,197,069	03/27/2007	Agazzi *et al.*	375	233	
	3V	7,206,418	04/17/2007	Yang *et al.*	381	92	
	3W	7,207,011	04/17/2007	Mulvey *et al.*	715	812	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	07/06/2012

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

403472863v1

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	4A	7,236,923	06/26/2007	Gupta	704	9	
	4B	7,277,854	10/02/2007	Bennett *et al.*	704	257	
	4C	7,305,381	12/04/2007	Poppink *et al.*	1	1	
	4D	7,328,155	02/05/2008	Endo *et al.*	704	251	
	4E	7,337,116	02/26/2008	Charlesworth *et al.*	704	254	
	4F	7,366,669	04/29/2008	Nishitani *et al.*	704	256	
	4G	7,415,414	08/19/2008	Azara *et al.*	704	270	
	4H	7,447,635	11/04/2008	Konopka *et al.*	704	275	
	4I	7,472,020	12/30/2008	Brulle-Drews	701	211	
	4J	7,472,060	12/30/2008	Gorin *et al.*	704	240	
	4K	7,493,259	02/17/2009	Jones *et al.*	704	257	
	4L	7,516,076	04/07/2009	Walker *et al.*	704	275	
	4M	7,536,297	05/19/2009	Byrd *et al.*	704	10	
	4N	7,536,374	05/19/2009	Au	706	55	
	4O	7,558,730	07/07/2009	Davis *et al.*	704	235	
	4P	7,574,362	08/11/2009	Walker *et al.*	704	275	
	4Q	7,577,244	08/18/2009	Taschereau	379	218.01	
	4R	7,606,708	10/20/2009	Hwang	704	257	
	4S	7,676,365	03/09/2010	Hwang *et al.*	704	240	
	4T	7,676,369	03/09/2010	Fujimoto *et al.*	704	270	
	4U	7,729,918	06/01/2010	Walker *et al.*	704	275	
	4V	7,788,084	08/31/2010	Brun *et al.*	704	7	
	4W	7,831,433	11/09/2010	Belvin *et al.*	704	275	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	07/06/2012

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	5A	7,873,523	01/18/2011	Potter *et al.*	704	275	
	5B	7,917,367	03/29/2011	Di Cristo *et al.*	704	270.1	
	5C	7,949,529	05/24/2011	Weider *et al.*	704	270	
	5D	7,949,537	05/24/2011	Walker *et al.*	704	275	
	5E	7,983,917	07/19/2011	Kennewick *et al.*	704	257	
	5F	8,015,006	09/06/2011	Kennewick *et al.*	704	236	
	5G	8,060,367	11/15/2011	Keaveney	704	247	
	5H	8,069,046	11/29/2011	Kennewick *et al.*	704	257	
	5I	8,073,681	12/06/2011	Baldwin *et al.*	704	9	
	5J	8,086,463	12/27/2011	Ativanichayaphong *et al.*	704	275	
	5K	8,112,275	02/07/2012	Kennewick *et al.*	704	240	
	5L	8,140,327	03/20/2012	Kennewick *et al.*	704	226	
	5M	8,140,335	03/20/2012	Kennewick *et al.*	704	257	
	5N	8,145,489	03/27/2012	Freeman *et al.*	704	257	
	5O	8,150,694	04/03/2012	Kennewick *et al.*	704	257	
	5P	8,155,962	04/10/2012	Kennewick *et al.*	704	257	
	5Q	2001/0049601	12/06/2001	Kroeker *et al.*	704	254	
	5R	2002/0015500	02/07/2002	Belt *et al.*	381	66	
	5S	2002/0035501	03/21/2002	Handel *et al.*	705	10	
	5T	2003/0016835	01/23/2003	Elko *et al.*	381	92	
	5U	2003/0110037	06/12/2003	Walker *et al.*	704	257	
	5V	2003/0115062	06/19/2003	Walker *et al.*	704	258	
	5W	2003/0120493	06/26/2003	Gupta	704	270.1	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	07/06/2012

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

INFORMATION DISCLOSURE CITATION IN AN APPLICATION	Docket Number	Application Number
	090611-0384800	12/705,311
	Applicant(s)	
	Robert A. KENNEWICK *et al.*	
	Filing Date	Group Art Unit
Form PTO-1449	February 12, 2010	2626

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE
	6A	2003/0144846	07/31/2003	Denenberg *et al.*	704	277	
	6B	2003/0158731	08/21/2003	Falcon *et al.*	704	231	
	6C	2003/0206640	11/06/2003	Malvar *et al.*	381	93	
	6D	2003/0212550	11/13/2003	Ubale	704	215	
	6E	2004/0098245	05/20/2004	Walker *et al.*	704	1	
	6F	2004/0167771	08/26/2004	Duan *et al.*	704	10	
	6G	2005/0137877	06/23/2005	Oesterling *et al.*	704	275	
	6H	2005/0143994	06/30/2005	Mori *et al.*	704	235	
	6I	2007/0033020	02/08/2007	(Kelleher) Francois *et al.*	704	226	
	6J	2007/0073544	03/29/2007	Millett *et al.*	704	277	
	6K	2007/0118357	05/24/2007	Kasravi *et al.*	704	10	
	6L	2008/0065386	03/13/2008	Cross *et al.*	704	270	
	6M	2008/0140385	06/12/2008	Mahajan *et al.*	704	9	
	6N	2009/0216540	08/27/2009	Tessel *et al.*	704	275	
	6O	2009/0271194	10/29/2009	Davis *et al.*	704	235	
	6P	2011/0112827	05/12/2011	Kennewick *et al.*	704	9	
	6Q	2011/0112921	05/12/2011	Kennewick *et al.*	705	26.1	
	6R	2011/0131036	06/02/2011	Di Cristo *et al.*	704	9	
	6S	2011/0131045	06/02/2011	Cristo *et al.*	704	249	
	6T	2011/0231182	09/22/2011	Weider *et al.*	704	9	
	6U	2011/0231188	09/22/2011	Kennewick *et al.*	704	236	
	6V	2012/0022857	01/26/2012	Baldwin *et al.*	704	9	
	6W	2012/0101809	04/26/2012	Kennewick *et al.*	704	9	
	6Y	2012/0101810	04/26/2012	Kennewick *et al.*	704	9	
	6Z	2012/0109753	05/03/2012	Kennewick *et al.*	705	14.58	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translations YES	Translations NO

OTHER DOCUMENTS *(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER	DATE CONSIDERED
/Fariba Sirjani/	07/06/2012

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.S./

403472863v1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 8125

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
12/705,311	02/12/2010	704	2626	090611-0384800

APPLICANTS

Robert A. Kennewick, Seattle, WA;
David Locke, Redmond, WA;
Michael R. Kennewick SR., Bellevue, WA;
Michael R. Kennewick JR., Bellevue, WA;
Richard Kennewick, Woodinville, WA;
Tom Freeman, Mercer Island, WA;
Stephen F. Elston, Seattle, WA;

**** CONTINUING DATA ***************************

This application is a CON of 10/618,633 07/15/2003 PAT 7,693,720
which claims benefit of 60/395,615 07/15/2002

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED **** ** SMALL ENTITY **
02/25/2010

Foreign Priority claimed ☐ Yes ☑ No 35 USC 119(a-d) conditions met ☐ Yes ☐ No Verified and Acknowledged /FARIBA SIRJANI/ Examiner's Signature	☐ Met after Allowance Initials	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
		WA	6	30	2

ADDRESS

Pillsbury Winthrop Shaw Pittman, LLP (NV)
PO Box 10500
McLean, VA 22102
UNITED STATES

TITLE

MOBILE SYSTEMS AND METHODS FOR RESPONDING TO NATURAL LANGUAGE SPEECH UTTERANCE

FILING FEE RECEIVED		
1142	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	12705311	KENNEWICK ET AL.
	Examiner	Art Unit
	FARIBA SIRJANI	2626

SEARCHED

Class	Subclass	Date	Examiner
704	270, 275	11-5-2011	fs

SEARCH NOTES

Search Notes	Date	Examiner
Inventor search.	11-5-2011	fs
Keyword search of PGPUB, USPAT and Derwent.	11-5-2011	fs
Search advice from M. Opsasnick and J. Wozniak.	11-7-2011	fs
Update inventor search and conduct keyword search.	7-24-2012	fs

INTERFERENCE SEARCH

Class	Subclass	Date	Examiner
704	270, 275, 251-257	7-24-2012	fs

/FARIBA SIRJANI/ Examiner.Art Unit 2626	

Google Exhibit 1019
Page 00996

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	12705311	KENNEWICK ET AL.
	Examiner	**Art Unit**
	FARIBA SIRJANI	2626

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	11/05/2011	07/24/2012							
	56	✓	✓							
	57	✓	✓							
	58	✓	✓							
	59	✓	✓							
	60	✓	✓							
	61	✓	✓							
	62	✓	✓							
	63	✓	✓							
	64	✓	✓							
	65	✓	✓							
	66	✓	✓							
	67	✓	✓							
	68	✓	✓							
	69	✓	✓							
	70	✓	✓							
	71	✓	✓							
	72	✓	✓							
	73	✓	✓							
	74	✓	✓							
	75	✓	✓							
	76	✓	✓							
	77	✓	✓							
	78	✓	✓							
	79	✓	✓							
	80	✓	✓							
	81	✓	✓							
	82	✓	✓							
	83	✓	✓							
	84	✓	✓							
	85	✓	✓							
	86		✓							
	87		✓							
	88		✓							
	89		✓							
	90		✓							
	91		✓							

Google Exhibit 1019
Page 00997

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	12705311	KENNEWICK ET AL.
	Examiner	**Art Unit**
	FARIBA SIRJANI	2626

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	11/05/2011	07/24/2012							
	92		✓							
	93		✓							
	94		✓							
	95		✓							
	96		✓							
	97		✓							
	98		✓							
	99		✓							

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L2	8649	704/270,275,251-257.ccls.	US-PGPUB; USPAT; DERWENT	OR	ON	2012/07/24 07:42
L4	0	"l3" and ((vehicle car telematics) and (part near3 order) and (cost price)).clm.	US-PGPUB; USPAT; DERWENT	OR	ON	2012/07/24 07:43
L5	32	"l3" and ((vehicle car telematics) and (cost price)).clm.	US-PGPUB; USPAT; DERWENT	OR	ON	2012/07/24 07:43
S1	2758	704/270.ccls.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:25
S2	35	Kennewick.in. near robert	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:25
S3	5	Kennewick.in. near michael	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:25
S4	17	Kennewick.in. near richard	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S5	91	locke.in. near david	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S6	19	freeman.in. near tom	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S7	12	elston.in. near stephen	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S8	1	S2 and S3 and S4 and S5 and S6 and S7	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:26
S9	2	"7693720".pn.	US-PGPUB; USPAT; DERWENT	OR	ON	2011/11/05 13:27

S10	9	("7693720").URPN.	USPAT	OR	ON	2011/11/05 13:27
S11	9	("7693720").URPN.	USPAT	OR	ON	2011/11/05 13:27
S12	187	("20010041980" | "20020049805" | "20020065568" | "20020082911" | "20020120609" | "20020124050" | "20020143535" | "20020188602" | "20020198714" | "20030014261" | "20030112267" | "20030182132" | "20040025115" | "20040044516" | "20040166832" | "20040199375" | "20050021334" | "20050021826" | "20050033574" | "20050114116" | "20050246174" | "20070033005" | "20070038436" | "20070050191" | "20080091406" | "20080103761" | "20080115163" | "20080133215" | "20090117885" | "20090144271" | "4430669" | "5155743" | "5274560" | "5377350" | "5386556" | "5424947" | "5471318" | "5475733" | "5499289" | "5500920" | "5517560" | "5533108" | "5537436" | "5539744" | "5557667" | "5563937" | "5590039" | "5617407" | "5633922" | "5675629" | "5696965" | "5708422" | "5721938" | "5722084" | "5742763" | "5752052" | "5754784" | "5761631" | "5774859" | "5794050" | "5797112" | "5799276" | "5802510" | "5832221" | "5878385" | "5878386" | "5892813" | "5895466" | "5902347" | "5911120" | "5918222" | "5926784" | "5933822" | "5953393" | "5963894" | "5963940" | "5987404" | "5991721" | "5995119" | "6009382" | "6014559" | "6021384" | "6044347" | "6049602" | "6058187" | "6078886" | "6081774" | "6101241" | "6119087" | "6134235" | "6144667" | "6160883" | "6173279" | "6175858" | "6185535" | "6192110" | "6192338" | "6195634" | "6195651" | "6208972" | "6219346" | "6219643" | "6233556" | "6233559" | "6246981" | "6272455" | "6292767" | "6314402" | "6366886" | "6381535" | "6385646" | "6393428" | "6404878" | "6408272" | "6411810").PN. OR ("6415257" | "6418210" | "6420975" | "6430285" | "6434523" | "6434524" | "6442522" | "6446114" | "6453153" | "6453292" | "6456711" | "6466654" | "6466899" | "6498797" | "6499013" | "6501833" | "6501834" | "6513006" | "6523061" | "6532444" | "6539348" | "6553372" | "6556970" | "6556973" | "6560576" | "6567778" | "6567797" | "6570555" | "6570964" | "6574597" | "6574624" | "6594257" | "6598018" | "6604077" | "6611692" | "6614773" | "6615172" | "6629066" | "6631346" | "6643620" | "6650747" | "6691151" | "6721001" | "6721706" | "6735592" | "6741931" | "6742021" | "6757718" | "6795808" |	US-PGPUB; USPAT; USOCR	OR	ON	2011/11/05 13:27