

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE**

VB ASSETS, LLC,

Plaintiff,

v.

AMAZON.COM SERVICES, LLC,

Defendant.

C.A. No. 24-_____

JURY TRIAL DEMANDED

COMPLAINT FOR PATENT INFRINGEMENT

This complaint addresses Amazon’s continuing infringement of Plaintiff’s patents through its use of natural language systems – including those products trained on large language models (LLMs) – for improved contextual conversations and voice driven advertisement and commerce.

Undaunted by a November 2023 jury verdict finding that Amazon willfully infringed four of Plaintiff’s patents and awarding an ongoing royalty, which (at the time of trial) equated to \$46.7 million in damages, Amazon continues to infringe not only the four previously adjudicated patents, but also five additional patents that are the subject of this complaint.

VoiceBox was a pioneer in conversational computing technology, developed the first smart speaker, and fostered a successful business with customers such as Toyota. Defendant Amazon.com Services LLC (“Defendant” or “Amazon”) desperately needed VoiceBox’s technology for its then top-secret Alexa project. Lured by the prospect of a partnership, VoiceBox disclosed its technology to Amazon, which unscrupulously crushed VoiceBox’s business and poached its engineers. VoiceBox filed a first lawsuit against Amazon in 2019. The jury unanimously concluded that Amazon willfully infringed VoiceBox’s patents. Yet, Amazon’s

willful infringement persists, compelling VoiceBox to file this second complaint against Amazon for willful patent infringement of U.S. Patent Nos. 10,297,249, 10,755,699, 11,087,385, 11,080,758, and 9,502,025 (collectively, the “Patents-in-Suit”).

FACTUAL BACKGROUND

A. 2002-2010: VoiceBox Invents Groundbreaking Voice Technology

1. VoiceBox pioneered voice-based search and commerce technology and invented what Amazon itself has described as “Echo-like” products long before Amazon.

2. In 2001, three brothers, Mike, Rich, and Bob Kennewick founded VoiceBox to bring natural language understanding (“NLU”) to a wide array of computer applications. They recognized that the typical computer speech-recognition systems forced human operators to adhere to a limited number of rigid speech prompts. These rigid prompts limited how systems were used and inhibited the widespread adoption of speech-recognition systems. The brothers believed that VoiceBox could become the first company to enable people to naturally and effectively interact with computer speech systems.

3. From its inception, VoiceBox engaged in intense research efforts to develop its NLU technology. As part of these efforts, VoiceBox achieved a significant milestone when it developed an early prototype called “Cybermind.” As demonstrated on the Seattle evening news, Cybermind was a voice-controlled speaker that could provide weather, recipes, sports scores, calendar updates, or play music.¹

¹<https://www.youtube.com/watch?v=DDcRyPnvWhw>



Figure 1: Cybermind Prototype

B. 2011: Amazon Steals Echo Concept from VoiceBox

4. In 2011, VoiceBox contacted Amazon to explore a potential business relationship where VoiceBox would provide core NLU services to Amazon. Amazon's corporate development department expressed interest and asked for "company and/or product overview slides" to facilitate an October 7, 2011 teleconference. In response, VoiceBox provided Amazon with a presentation that described its award-winning patented technology and explicitly referred to VoiceBox's "patented Contextual Speech Technology." Slides from that presentation are reproduced below:

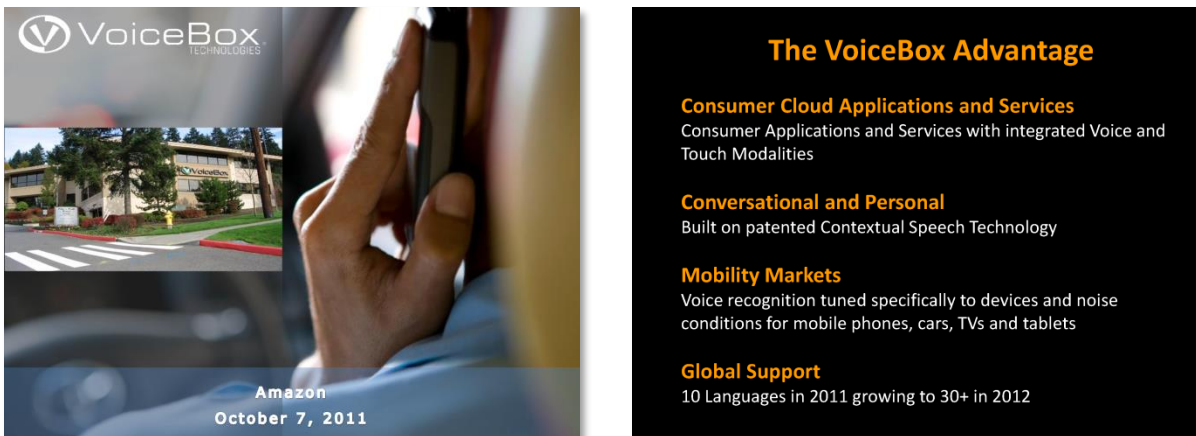


Figure 2: 2011 Slides

5. Amazon was so impressed by the technology VoiceBox presented on October 7, 2011, that its representative emailed VoiceBox the next business day to invite VoiceBox to visit Amazon's offices on October 19, 2011. That meeting was with Douglas Booms, Amazon's Vice President of Worldwide Corporate Development, as well as engineers and product/business development members of Amazon's devices and digital teams. The email from Amazon stated that this was "the right audience to discuss [VoiceBox's] personal digital assistant and underlying conversational voice technology." On information and belief, in addition to Mr. Booms and several engineers, at least the following Amazon executives attended the meeting: Nick Komorous (Director, Corporate Development), Ian Freed (VP, Amazon Devices), Greg Hart (VP, Digital), Al Lindsay (Director, Software Engineer), and Frederic Deramat (Senior Principal Engineer).

6. Two days later, Amazon's Mr. Komorous emailed VoiceBox and asked to visit the company's office for a "deeper dive." Mr. Komorous wanted to meet as soon as possible. VoiceBox agreed to meet at VoiceBox's office on October 26, 2011.

7. In advance of the meeting, Mr. Komorous sent a detailed set of technical questions that would help Amazon's "tech team understand the scope and [r]ange of things [Amazon] can try to tackle with VoiceBox[] as a partner." Mr. Komorous also conveyed that Amazon's culture was "engineering heavy" and asked that VoiceBox "have engineering and speech representation at the meeting."

8. The October 26, 2011 meeting at VoiceBox's office ran from 10 AM until 12 PM with some Amazon engineers staying until around 2:30 PM. The meeting included a tour with technology demonstrations, a review of the technical architecture for VoiceBox's server software, and a disclosure of plans for next generation products.

9. During the meeting, VoiceBox made a slide presentation providing even more detail about VoiceBox's patented technology and informed Amazon that VoiceBox had 12 patents at the time with an additional 14 pending applications. The slide deck included the following slides regarding the VoiceBox patent portfolio.

VoiceBox Patents

- VoiceBox has been awarded 12 patents for contextual and conversational speech technologies
- VoiceBox has an additional 14 pending applications
- Our most recent patent (at right) for processing natural language with context specific domain agents also incorporates personification

United States Patent
Kennewick et al.

(10) Patent No.: **US 8,015,006 B2**
(45) Date of Patent: **Sep. 6, 2011**

(54) **SYSTEMS AND METHODS FOR PROCESSING NATURAL LANGUAGE SPEECH UTTERANCES WITH CONTEXT-SPECIFIC DOMAIN AGENTS**

(75) Inventors: Robert A. Kennewick, Seattle, WA (US); David Locke, Redmond, WA (US); Michael R. Kennewick, Sr., Bellevue, WA (US); Michael R. Kennewick, Jr., Bellevue, WA (US); Richard Kennewick, Woodinville, WA (US); Tom Freeman, Mercer Island, WA (US)

(73) Assignee: VoiceBox Technologies, Inc., Bellevue, WA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 617 days.

(21) Appl. No.: 12/136,797
(22) Filed: May 30, 2008

(65) **Prior Publication Data**
US 2008/0235023 A1 Sep. 25, 2008

Related U.S. Application Data
(63) Continuation of application No. 10/452,147, filed on Jun. 3, 2005, now Pat. No. 7,306,209.
(60) Provisional application No. 60/384,388, filed on Jun. 3, 2002.

(51) Int. Cl. (2006.01)
G06F 15/28 (2006.01)
G06F 15/22 (2006.01)
G06F 17/00 (2006.01)
G06F 17/27 (2006.01)
G06F 17/30 (2006.01)

(52) U.S. Cl. 704/236; 704/244; 704/246; 704/257; 707/709; 707/723; 707/771

(38) Field of Classification Search 704/256; 704/246; 257; 270; 244; 707/709; 723; 725; 707/771
See application file for complete search history.

(56) **References Cited**
U.S. PATENT DOCUMENTS
4,430,689 A 2/1984 Chung 358/122 (Continued)
FOREIGN PATENT DOCUMENTS
EP 1,320,045 A2 6/2003 (Continued)
OTHER PUBLICATIONS
Bosters, "How to Enable Hands Drivers to Talk to Cars", Charles Schuch & Co., Inc., Jul. 28, 2003, 1 page. (Continued)
Primary Examiner—Marta Lerner
(74) Attorney, Agent, or Firm—Pillsbury Winthrop Shaw Pittman LLP

(57) **ABSTRACT**
Systems and methods for receiving natural language queries and/or commands and executing the queries and/or commands. The systems and methods overcome the deficiencies of prior art speech query and response systems through the application of a complete speech-based information query, retrieval, presentation and command environment. This environment makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands in multiple domains. Through this integrated approach, a complete speech-based natural language query and response environment can be created. The systems and methods create, store and use extensive personal profile information for each user, thereby improving the reliability of determining the context and presenting the expected results for a particular query or command.

11 Claims, 7 Drawing Sheets

Figure 3: Slide from 2011 Presentation

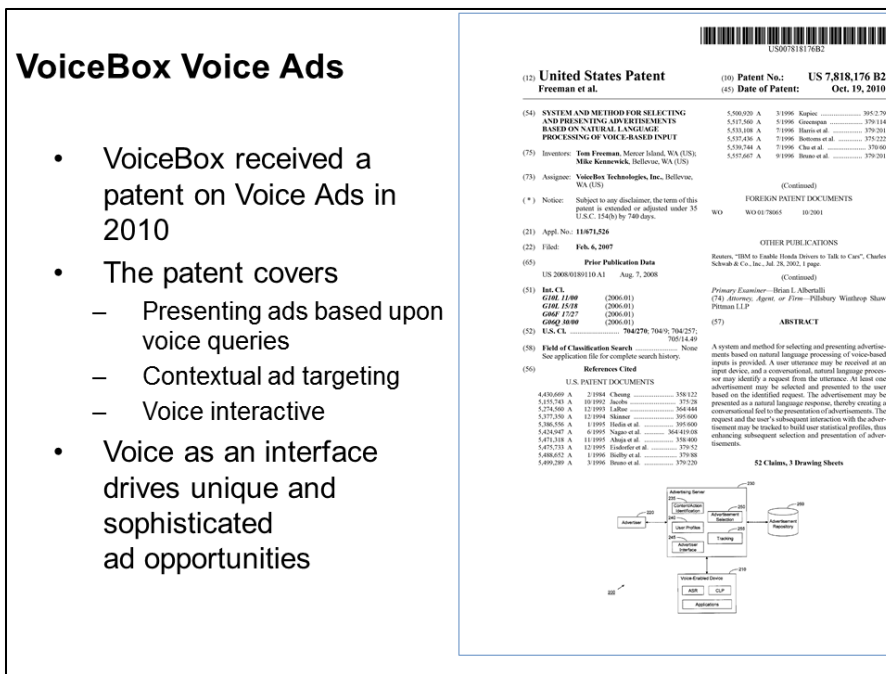


Figure 4: Slide from 2011 Presentation

10. VoiceBox's presentation also proposed a business arrangement where VoiceBox would provide "Voice Services" to Amazon. The VoiceBox Voice Services from the presentation are shown below:



Figure 5: Slide from 2011 Presentation

11. On information and belief, some of the Amazon personnel involved in the 2011 meetings became technical leaders for Amazon's Alexa Products² while others became high-level executives with close working relationships with the senior leadership of Amazon. For instance, on information and belief, Mr. Freed was the Technical Advisor to the CEO, effectively Jeff Bezos's "shadow" advisor. On information and belief, Mr. Freed later became Vice President, Amazon Devices where he led a team of engineers working on Amazon's Alexa Products. On information and belief, Mr. Hart has also been Mr. Bezos's "shadow" advisor—he held the Technical Advisor to the CEO position around 2011. On information and belief, around 2011, Mr. Lindsay was promoted to Vice President managing the Alexa Engine Software team. On information and belief, around 2011, Mr. Deramat was promoted to the position of Vice President & Distinguished Engineer for Amazon Alexa. On information and belief, around 2011, Mr. Thimsen was promoted to the position of Director of Engineering for Amazon Echo. As for Mr. Typrin, he states on his LinkedIn page that he is "[o]ne of the founding members of the team that shaped the vision and direction for Amazon's Echo and Alexa Voice Services."

12. On information and belief, Amazon did not have its own NLU technology, nor did Amazon have NLU expertise in-house. As a result, Amazon had to look externally to provide that NLU technology and expertise as it was developing the Alexa Products. This was Amazon's true motive for meeting with VoiceBox.

13. A couple of days after the last meeting in 2011, VoiceBox sent an email to Amazon asking for Amazon's feedback. Mr. Komorous from Amazon replied that Amazon was "still discussing internally how contextual speech / cybermind 2012 could play a part in

² "Alexa Products" refers to Alexa, including but not limited to all associated hardware and software, Echo devices, Alexa mobile apps, Fire TV, Amazon smart glasses, and third-party devices with Alexa integrated.

[Amazon's] future.” Amazon abruptly went silent. Amazon did not provide the results of these discussions and did not pursue a business relationship with VoiceBox.

C. 2011-2016: VoiceBox Gains Industry Recognition for Its Technology

14. VoiceBox was obviously disappointed with Amazon's conduct, but moved past its disappointment focusing on its successful business. For example, after learning about VoiceBox's technology, Toyota hired it to build a sophisticated NLU speech interface for its Lexus automobiles. Toyota was so impressed with VoiceBox's technology that it expanded the relationship such that VoiceBox provided voice and NLU capability for Toyota's award-winning Entune multimedia system.

15. Additional customers included TomTom, Pioneer, Chrysler, Dodge, and Magellan. VoiceBox had software applications that ran on smart speakers, in-car systems, smartphones, smart TVs, computers, tablets, e-readers, and personal navigation devices. A few examples are illustrated below:

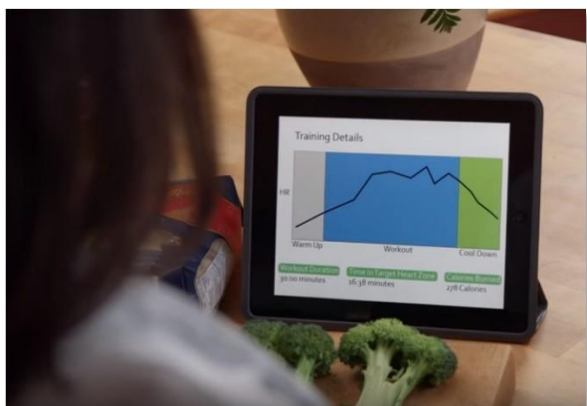


Figure 6: Connected Home



Figure 7: Mobile

16. In 2013, the Institute of Electrical and Electronics Engineers (“IEEE”) ranked VoiceBox number 13 in patent power for the computer software industry. VoiceBox had become

the leader in conversational Artificial Intelligence (“AI”), including Voice Recognition (VR), NLU, and AI services.

17. As illustrated in the following company photograph, VoiceBox was, at this time, optimistic about its technology and its future.



D. 2014-2017: Amazon Launches Alexa and Poaches VoiceBox Employees

18. Everything changed in November 2014. Amazon announced the launch of Alexa, a virtual assistant, along with the first-generation Echo product, a smart speaker. VoiceBox was stunned and dismayed. Amazon had clearly mimicked VoiceBox’s products and technology. The 2011 meetings were a ploy that Amazon used to learn as much as it could about VoiceBox’s technology, business, products and vision for the future.

19. Amazon’s efforts to supplant and replace VoiceBox did not stop with mimicry. In 2016, Amazon abruptly hired Philippe Di Cristo, who was VoiceBox’s Chief Scientist. While at VoiceBox, Dr. Di Cristo gained knowledge of the company’s voice technology and had full access to VoiceBox’s intellectual property. As Dr. Di Cristo explains on his LinkedIn Page, he had worked on an “Amazon Echo-like system” while at VoiceBox.

20. On information and belief, Dr. Di Cristo was pivotal in soliciting additional VoiceBox employees to join Amazon. For example, on January 10, 2017, Amazon rented the

entire high-end seafood restaurant near VoiceBox to host an “Evening with the Leadership of Amazon Voice & Advanced Shopping,” which Amazon expressly described as an “invite only networking event for Voice Box employees . . . to talk . . . about opportunities at Amazon.” Dr. Di Cristo’s boss invited a large number of VoiceBox engineers—including Mike Kennewick’s own son—and shared that Amazon was “building a world-class speech & NLU engineering team” and “[y]our profile looks quite relevant and we’d love to talk to you and see if there’s a fit.” Dr. Di Cristo was originally scheduled to speak at the event—but abruptly withdrew.

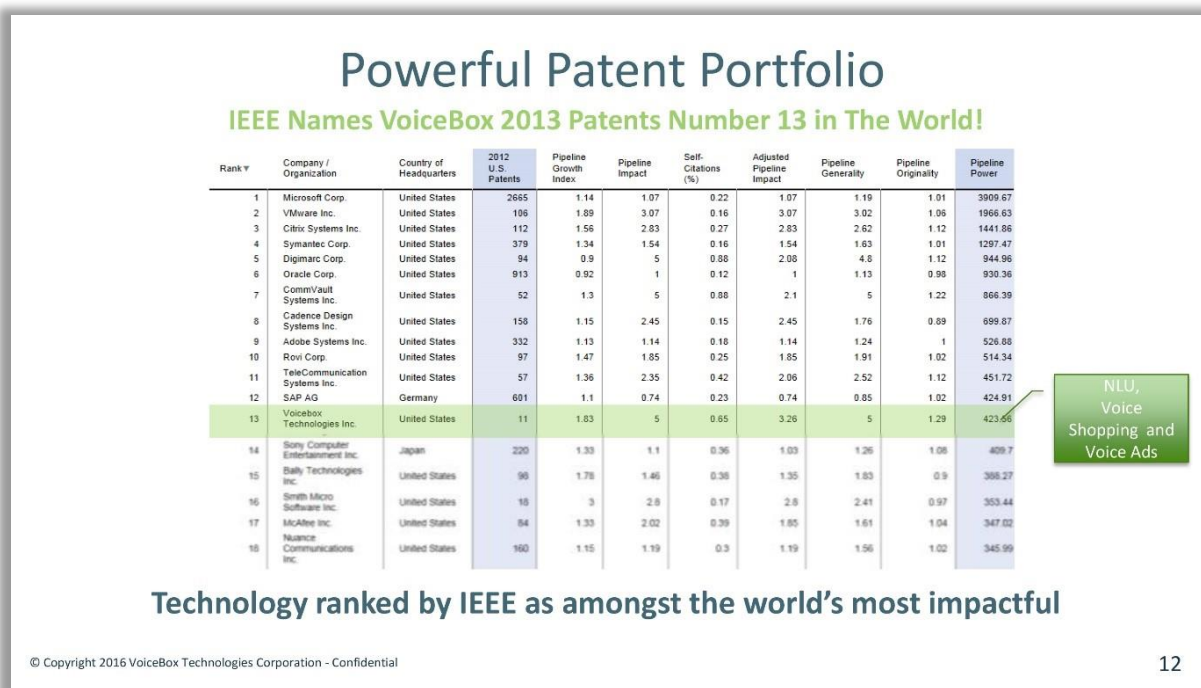
21. By January 17, 2017, it was apparent that Amazon was using its vast resources to recruit and capture VoiceBox’s engineering talent. In response, Mike Kennewick, CEO of VoiceBox, sent a letter to Jeff Bezos, CEO of Amazon, to propose a business solution. The letter explains that, at the time, VoiceBox had “a deep portfolio of technology and IP, including a large number of significant patents not only in NLU but also in Voice[]Commerce, running on over 200 million devices.”

E. 2017: Amazon Steals VoiceBox’s Automotive Business

22. Mr. Booms, Amazon’s Vice President of Worldwide Corporate Development, responded the very next day. He requested a meeting at VoiceBox’s offices for purposes of “go[ing] fairly deep on the technology, data, customer relationships.”

23. On February 2, 2017, the parties met at VoiceBox’s office. Amazon came with a team of technologists from its Alexa program, including on information and belief, Manoj Sindhvani (Director, Alexa), Karthik Ramakrishnan (Senior Manager, Alexa software), Nikko Strom (Scientist, Alexa), and Deepesh Mohnani (Alexa Voice Services Product Management) joined by Mr. Booms. VoiceBox provided a detailed technical presentation, which included information about patents and pending applications then owned by VoiceBox.

24. Slides in the presentation listed all VoiceBox patents and published applications at the time, including U.S. Patent Publication No. 2015/0228276, later issued as the '249 patent.



The slides further confirmed the strength of VoiceBox's patent portfolio in the areas of NLU, Voice Shopping and Voice Ad technology. The presentation also emphasized industry recognition of VoiceBox's patent portfolio:

25. Following the meeting, Mr. Booms emailed VoiceBox to request even more technical details. Then, on February 20, 2017, Amazon specifically asked VoiceBox for a list of all patents owned by the company.

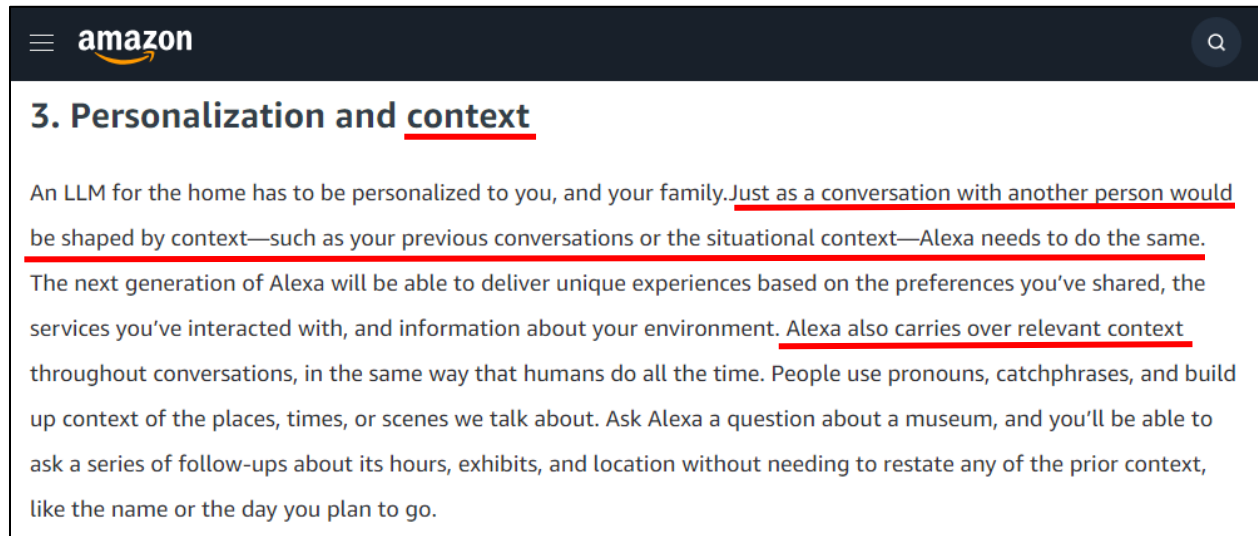
29. At this point, for the first time, Mike Kennewick appreciated that Amazon was targeting VoiceBox and that his company would not survive long term. Mr. Kennewick began looking for buyers. Unfortunately, most of VoiceBox's value had already been eviscerated by Amazon. The best Mr. Kennewick could do was to sell VoiceBox to Nuance Communications, Inc. ("Nuance"), a computer software technology company, for a fraction of the company's true value. Despite the sale to Nuance, Mr. Kennewick managed to hold onto key VoiceBox patents even after the sale to Nuance to hold Amazon accountable for its misconduct.

F. 2019-2023: Amazon Found to Willfully Infringe

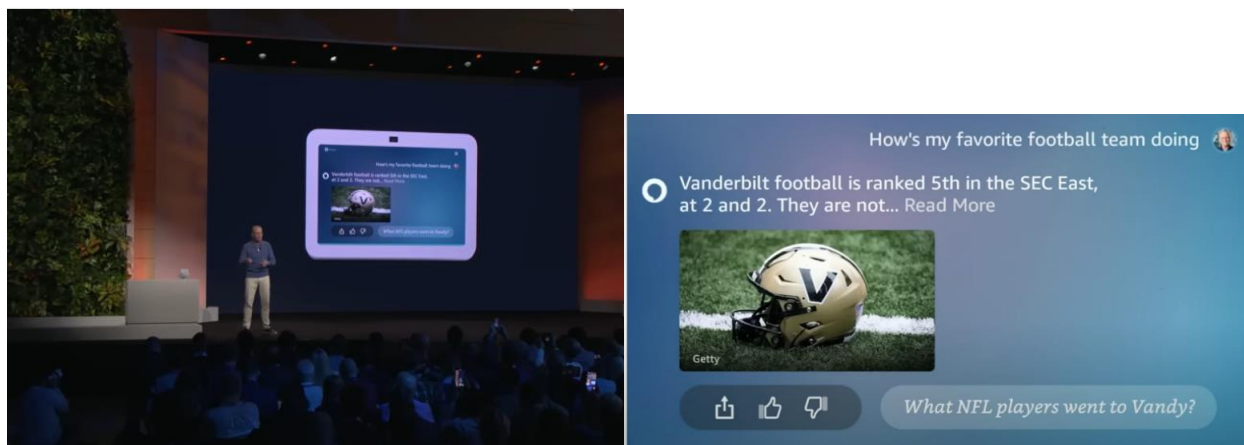
30. On July 29, 2019, VoiceBox filed a complaint for patent infringement against Amazon in this District, asserting infringement of certain patents related to and in the same patent families as the Patents-in-Suit.

31. Amazon pulled out all the stops in litigation, including challenging the patent eligibility of each patent, putting up obstinate witnesses who went so far as disputing the veracity of their own internal documents, and submitting a 1500-page opening expert report on alleged invalidity. At trial, Amazon abandoned nearly all of those many invalidity arguments and focused on a simple theme: Alexa was a neural network-based system that could not infringe VoiceBox patents. After hearing all the evidence, the jury rejected Amazon's theme as utterly false. On November 8, 2023, a Delaware federal jury found that Amazon willfully infringed four VoiceBox patents (related to the Patents-in-Suit) and determined that Amazon should pay VoiceBox an ongoing royalty for its infringement which, at the time of trial, equated to \$46.7 million in damages. The jury also resoundingly rejected Amazon's invalidity defenses.

32. Despite this history, Amazon doubled down on its exploitation of VoiceBox's technology. In September 2023, shortly before trial commenced, Amazon heavily promoted Alexa's improved conversation capabilities based on personalization and context:



33. Amazon's announcement came while it was, in effect, arguing in court that Alexa did not utilize "context" at all. Amazon accompanied its announcement with a live show-and-tell led by Dave Limp, one of Alexa's highest-ranking executives. Mr. Limp demonstrated an example where Alexa knew the identity of his favorite football team based on his previous conversations and provided a response tailored to his prior interactions:



34. Following trial, Amazon moved still closer to VoiceBox, introducing a smart speaker nearly identical in appearance to the original VoiceBox.



35. On May 1, 2024, VoiceBox sent Amazon a letter notifying it of what Amazon already knew. Namely, that Alexa Products infringed the patents-in-suit. Amazon denied infringement, but its reasons, as before, were meritless.

THE PARTIES

36. VB Assets, LLC, is a limited liability company organized under the laws of Delaware and has its principal place of business at 1229 120th Ave. NE, Suite A, Bellevue, WA 98005.

37. On information and belief, Amazon.com Services LLC is a limited liability company organized under the laws of Delaware and has a principal place of business at 410 Terry Ave. N, Seattle, WA 98109. On information and belief, Amazon.com Services LLC has participated in the sale or offer for sale of one or more Alexa Products and/or has provided digital services and content for use by Alexa Products.

JURISDICTION AND VENUE

38. This Court has original jurisdiction over the subject matter of this action under 28 U.S.C. §§ 1331 and 1338(a) because the action arises under the patent laws of the United States.

39. Personal jurisdiction over Amazon is proper in this District based on one or more of the following: its presence in this judicial district; it has availed itself of the rights and benefits of the laws of Delaware; or it has derived substantial revenue from sales of Alexa Products in Delaware, and it has systematic and continuous business contacts with Delaware. Amazon was incorporated in Delaware and/or formed under the laws of Delaware and Amazon designs Alexa Products, which are advertised, offered for sale, sold, and used in Delaware.

40. Venue is proper in this district under 28 U.S.C. § 1400(b) and 28 U.S.C. §§ 1391(b)(1), (b)(2). For purposes of § 1400(b), Amazon was incorporated in Delaware and/or formed under the laws of Delaware and therefore resides within this District. For purposes of § 1391(b)(1), (b)(2), Amazon resides in the District of Delaware by virtue of being incorporated in Delaware and/or formed under the laws of Delaware.

OVERVIEW OF THE PATENTS-IN-SUIT

The '249 and '699 Patents

41. United States Patent Number 10,297,249 (“the '249 patent”), entitled “System and Method for a Cooperative Conversational Voice User Interface,” was duly and legally issued on May 21, 2019, and names Larry Baldwin, Tom Freeman, Michael Tjalve, Blane Ebersold, and Chris Weider as the inventors. Attached as Exhibit A is a true and correct copy of the '249 patent.

42. The '249 patent claims, among other things, a system for facilitating natural language system responses via short-term knowledge generated based on one or more prior multi-modal device interactions: one or more physical processors programmed with one or more computer program instructions which, when executed, cause the one or more physical processors to: receive, during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance; receive a second voice input comprising the

first natural language utterance via a second input device; compare the first voice input with the second voice input; filter sound from the first voice input and the second voice input based on the comparison; obtain, during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input; generate the short-term knowledge based on at least the first voice input and the first non-voice input; determine, based on the short-term knowledge, a first context for the first natural language utterance; determine, based on the first context, an interpretation of the first natural language utterance; and generate, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.

43. VB Assets, LLC is the assignee of the entire right, title, and interest in the '249 patent.

44. United States Patent Number 10,755,699 ("the '699 patent"), entitled "System and Method for a Cooperative Conversational Voice User Interface," was duly and legally issued on August 25, 2020, and names Larry Baldwin, Tom Freeman, Michael Tjalve, Blane Ebersold, and Chris Weider as the inventors. Attached as Exhibit B is a true and correct copy of the '699 patent.

45. The '699 patent claims, among other things, a system for generating natural language system responses adapted based on a user's manner of speaking, the system comprising: one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to: receive a user input comprising a natural language utterance; recognize one or more words or phrases from the natural language utterance; identify a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance; determine an

interpretation of the natural language utterance based on the identified context; accumulate short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system; accumulate long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period; identify a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and generate a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.

46. VB Assets, LLC is the assignee of the entire right, title, and interest in the '699 patent.

47. Voice user interface systems in existence before the inventions of the '249 and '699 patents were typically of the "Command and Control" type. Such systems used verbal menus to restrict information that a person can provide at a given point. For example, the voice system would present the list of available options either verbally and/or on a screen. The user could then respond by speaking the menu item. Such a system could include numerous menus that the user would have to get through in order to convey the desired information to the system and/or to cause the system to take the desired action. This type of system often fails to provide a seamless conversational experience.

48. The inventors of the '249 and '699 patents recognized a significant problem with the Command and Control systems in that users would have to memorize exact words and phrases in order to interact with the system. This required significant learning because the user

had to know which words and phrases to use in order to make a request of a Command and Control voice user interface system. Additionally, the process of stepping through menus could be time-consuming and, in some cases, might dissuade a user from utilizing the voice-based system.

49. To overcome the shortcoming of prior art systems, the inventors provided a solution that used an “Automatic Speech Recognizer” (or ASR) to generate a preliminary interpretation and provide that preliminary interpretation to a “conversational speech engine” for further processing. The conversational speech engine—which was not well-understood, routine, or conventional—could be implemented locally on a user device or remotely on a server. In certain embodiments, the conversational speech engine included a conversational language processor, voice search engine / free form voice search module with context domain agents, and a context determination process. The conversational speech engine communicates with databases to generate an adaptive conversational response.

50. Through the use of a conversational speech engine, the ’249 and ’699 patents advantageously rely on conversational responses which, in some embodiments, use short-term, and in some embodiments long-term shared knowledge, about user utterances to determine a context for the request, infer additional information about a request, and provide an adaptive conversational response. For example, Figure 1 shows an architecture for the conversational speech engine of the cooperative conversational voice user interface.

51. The inventors were thereby able to improve the functioning of voice user interface systems which improved the operation of those systems in an unconventional manner. For example, the innovations in the ’249 and ’699 patents allowed a user to converse naturally with a voice user interface system instead of “dumbing down” their requests to match the simple sets of

instructions that existing Command and Control systems required. In this regard, one of the problems faced by the inventors was necessarily rooted in voice user interface technology specifically arising in the realm of voice user interface systems. Indeed, the commercial success and industry accolades provide objective evidence as to VoiceBox's innovative approach through the use of unconventional technology.

52. The '249 patent further describes and claims a system for facilitating natural language system responses using short-term knowledge generated based on multi-modal device interactions including a system that can receive and process inputs from multiple modalities, such as voice and non-voice inputs, to generate short-term knowledge. This allows for a more comprehensive understanding of user interactions by integrating different types of input data. The system also includes mechanisms to compare multiple voice inputs and filter out noise, enhancing the accuracy of speech recognition. This ensures that the system can handle variations in voice input quality and environmental noise effectively.

53. The known prior art in the field of voice user interfaces neither taught accumulating short-term knowledge nor expressed any appreciation for the substantial advantages associated with utilizing this shared knowledge for various purposes in a conversational speech engine. Such uses include to identify context, infer additional information about an utterance that contains insufficient information to complete a request, establish an intended meaning for an utterance within the context based on the additional information inferred about the utterance, and generate a response based on the intended meaning established within the identified context. In this regard, accumulating and using both short-term and long-term knowledge was not well-understood, routine, or conventional and stands in sharp contrast to

the conventional and routine approach of Command and Control systems that require a user to use rigid menus to establish context before making a request.

54. Additionally, the known prior art in the field of voice user interfaces did not teach receiving multiple voice inputs and filtering sounds from a first voice input and second voice input, whereby the filtering is based on a comparison. Such comparison helps in identifying and discarding inconsistent or extraneous sounds that do not match across the inputs. The system can also compare speech signals from the various voice inputs to filter out irrelevant noise and focus on the primary voice input. The context determination process can also filter out words that do not fit into the identified context.

55. During prosecution of the '249 patent, the examiner rejected numerous application claims as being unpatentable over Byers (US 6,219,645), in combination with other prior art, such as Kennewick (US 2004/0044516), Ehlen (US 2004/0006475), and Perkins (US 7,072,888). The USPTO eventually granted the claims of the '249 patent finding that the prior art (Byers) did not describe "where 'sound from the first voice input and the second voice input' (i.e. sound that is based on both voice inputs from the first and second input devices) are filtered, where the filtering is based on the comparison." This was not well-understood, conventional, or routine to a person of ordinary skill in the art.

56. The '699 patent further enhances voice user interfaces by using short-term knowledge accumulated based on one or more natural language utterances received during a predetermined time period. This accumulation occurs during a conversation between the user and the computer system. The system is designed to model human conversations, where short-term session data is expired after a psychologically appropriate amount of time to humanize system behavior and reduce the likelihood of contextual confusion based on stale data. Consequently,

the invention significantly improves conversational accuracy and the natural flow of interactions between the user and the computer system.

57. The '699 patent further provides for a system that can identify a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge. In some embodiments, the manner includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.

58. During prosecution of the '699 patent, the examiner rejected numerous application claims as being unpatentable over Kargman (US 2005/0015256), Morin (US 5,748,841), Kennewick (US 2004/0044516), Perkins (7,072,888), and/or several other prior art references. Based, at least in part, on the inventors' amendments to the claims, the USPTO eventually granted the claims of the '699 patent finding them novel and non-obvious over the prior art. As indicated, a key distinction—which the examiner eventually agreed with—is that Kargman did not teach or suggest that “context is determined based on short-term knowledge and long-term knowledge (i.e. intended meaning is not the same as context).” Moreover, “[n]either Morin nor Perkins teach/suggest where short-term knowledge that has been expired is included in long-term information (expired information is simply deleted in Morin and Perkins), and Kennewick does not specifically teach that both short-term and long-term information is used to determine context (even though dialog history and other examples in Kennewick are likely to be based on utterances entered long enough ago to be considered ‘long-term’).” Nor did they teach “expiring short-term knowledge that is based on natural language utterances received prior to the predetermined time period being expired and added to long-term storage.” None of this was well-understood, conventional, or routine to a person of ordinary skill in the art.

The '385 Patent

59. United States Patent Number 11,087,385 (“the '385 patent”), entitled “Voice Commerce,” was duly and legally issued on August 10, 2021, and names Michael R. Kennewick, Sr., as the inventor. Attached as Exhibit C is a true and correct copy of the '385 patent.

60. The '385 patent claims, among other things, a system for providing voice commerce, the system comprising: one or more physical processors programmed with computer program instructions which, when executed, cause the one or more physical processors to: receive a single first user input comprising a natural language utterance; provide the natural language utterance as an input to a speech recognition engine; obtain one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine; search one or more databases of products or services based on the one or more words or phrases; select, without further user input other than the single first user input, a product or service from the database to be purchased based on the search; receive a second user input indicating confirmation by a user to complete a purchase transaction of the selected product or service; and complete, without further user input after the receipt of the second user input, a purchase transaction of the selected product or service.

61. VB Assets, LLC is the assignee of the entire right, title, and interest in the '385 patent.

62. Online shopping systems in existence before the inventions of the '385 patent typically required a user to browse a website to locate a product, make payment, and have the product delivered.

63. The inventor recognized a significant problem with such systems in that a user must search a website in order to locate a product or service to be purchased and fill-out

numerous payment and shipping forms to complete checkout. This problem was exacerbated on a mobile electronic device because such devices typically have small screens and keyboards making it hard for the user to search for the product or service to purchase and input payment and shipping information.

64. In certain embodiments, the '385 patent advantageously provides a voice commerce system with a voice user interface for online shopping. For example, Figure 1 shows an architecture for the voice commerce system. The voice commerce system includes user input processing instructions 122, which may comprise a speech recognition engine and a natural language processing engine. Figure 2 shows a system for facilitating natural language processing for the voice commerce system including a speech recognition engine and a natural language processing engine.

65. The '385 patent claims and describes a system that advantageously determines a product or service to be purchased on behalf of the user based on a natural language utterance. To do so, the system receives a natural language utterance and uses a speech recognition engine to recognize words and phrases from the natural language utterance. An example of this speech recognition engine is shown in Figure 2 as the Speech Recognition Engine 220. From the words and phrases the system identifies a context and determines a product or service to be purchased without further user input identifying a product or service. Exemplary components for performing this natural language processing are the Natural Language Processing Engine 230 shown in Figure 2 and the Transaction Preparation Instructions 124 shown in Figure 1.

66. By providing this innovation, the inventor was able to improve the functioning of voice user interfaces for online shopping systems thereby improving the operation of those systems in an unconventional manner. For example, the innovations in the '385 patent determine

a product or service to be purchased based on a natural language utterance. This distinguishes the '385 patent from existing online shopping systems that required a user to search a website to locate a product or service to be purchased. In this regard, one of the problems faced by the inventors was necessarily rooted in online shopping technology specifically arising in the realm of online shopping.

67. The system has a further advantage in that it prepares and completes a transaction without further user input identifying a product or service, payment information, and/or shipping information. The system may obtain payment information with which to pay for the product or service without further user input identifying payment information. The system obtains shipping information with which to deliver the product or service without further user input identifying shipping information. Then the system completes a purchase transaction for the product or service without further user input identifying a product or service, payment information, and/or shipping information. An exemplary component for performing the above is the checkout management instructions 126 shown in Figure 1 of the '385 patent.

68. In doing so, the inventor was able to improve the functioning of online shopping and voice user interface systems thereby improving the operation of those systems in an unconventional manner. For example, the innovations in the '385 patent prepare and complete a transaction without further user input identifying a product or service, payment information, or shipping information. This distinguishes the '385 patent from existing online shopping systems that required a user to search for a product or service and fill-out numerous payment and shipping forms to complete checkout. In this regard, one of the problems faced by the inventors was necessarily rooted in online shopping technology specifically arising in the realm of online shopping.

69. During prosecution of the '385 patent, the examiner rejected numerous application claims as being unpatentable over Cohen (US 6,859,776) and Kinsey (US 2014/0136259). In response, VoiceBox argued that “Cohen fails to teach or suggest searching a database of products or services – it simply discloses stored words and dictionaries. Further the latter portion cited by the Examiner, Cohen discloses a series of quires and answers used to, for example, select a flight, that is quite different from the claimed invention. The claimed invention selects a product or service from a database based on *single user input* (the claimed ‘single first user input’), without the further user input.” VoiceBox went on to explain that the combination of references lacked numerous claim limitations. The USPTO eventually granted the claims of the '385 patent finding them novel and non-obvious over Cohen, Kinsey, and other cited references.

70. Conventional wisdom in the context of shopping was keenly focused on menu-based systems. Indeed, web-stores were (and still are) based on product categorizations. The typical user experience involves going through a series of menus to narrow down the particular product. As such, there was particular focus and motivation to emulate the menus in a speech-based system. That is, for the computer to read options which the end-user selects before moving to the next menu. The '385 patent represents a dramatic departure by “selecting...without further user input other than the single first user input, a product or service from the database to be purchased based on the search....”

71. Additionally, conventional wisdom was that to complete an online purchase the user would either have to provide a shipping address or, at minimum, affirmatively select a predefined address. While this approach was perhaps feasible in the context of a visual user interface, the inventor recognized that it was an unnecessary and distracting step in the context of voice-controlled purchase transactions. Therefore, the '385 patent requires

“completing...without further user input after the receipt of the second user input, a purchase transaction of the selected product or service.” In some embodiments, the ’385 patent further requires “obtaining...shipping information with which to deliver the selected product or service, wherein the shipping information specifies a name or address of a recipient to which the selected product or service is to be delivered after the selected product or service is purchased, and wherein the purchase transaction is completed based on the shipping information without receiving confirmation of the shipping information by the user.”

72. Additionally, “obtaining ... a predetermined set of sellers specified by an administrator of the system that is different than the user” was not well-understood, routine, or conventional. Indeed, such information could be used to reduce the amount of information manually entered (or selected) by the user which reduces the complexity associated with using voice for commerce. This was an unconventional approach to solving the problems associated with transactions via voice.

73. A skilled artisan would not consider the claim limitations of the ’385 patent – either alone or in combination – to recite well-understood, routine, or conventional concepts. Instead, a person of ordinary skill in the art would recognize that the claim limitations of the ’385 patent are directed to the inventive concepts described in the specification and prosecution history.

The ’758 Patent

74. United States Patent Number 11,080,758 (“the ’758 patent”), entitled “System and Method for Delivering Targeted Advertisements and/or Providing Natural Language Processing Based on Advertisements,” was duly and legally issued on August 3, 2021, and names Tom

Freeman and Mike Kennewick as the inventors. Attached as Exhibit D is a true and correct copy of the '758 patent.

75. The '758 patent claims, among other things, a system for processing natural language utterances that include selecting and presenting purchase opportunities based thereon, the system comprising: one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to: provide a natural language utterance as an input to a speech recognition engine; receive words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine; determine a context for the natural language utterance based on the recognized words or phrases; select a purchase opportunity based on the determined context; deliver the selected purchase opportunity via an electronic device in communication with the one or more processors; track an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity, wherein the subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity; build or update a user-specific profile based on the tracked interaction pattern; and interpret a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.

76. VB Assets, LLC is the assignee of the entire right, title, and interest in the '758 patent.

77. Before the invention of the '758 patent, voice user interface systems were typically difficult to use, in part, because they had complex human to machine interfaces. Such

systems forced a user to navigate through a series of menus and provide a series of user inputs to perform a relatively simple task.

78. The inventors recognized significant problems with existing systems. The systems did not allow a user to directly issue a request without having to memorize specific syntaxes, words, phrases, concepts, semantic indicators, or other keywords/qualifiers. Similarly, when users were uncertain of particular needs, many existing systems did not engage the user in a productive, cooperative dialogue to resolve requests and advance a conversation. Instead, many existing speech interfaces forced users to use a fixed set commands or keywords to communicate requests in ways that systems could understand. Using existing voice user interfaces, there was virtually no option for dialogue between the user and the system to satisfy mutual goals.

79. The inventors recognized other problems with existing systems. The lack of adequate voice user interfaces resulted in missed opportunities for providing valuable and relevant information to users. Not only did this potentially leave user requests unresolved, in certain instances, providers of goods and services may have lost 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 left a large segment of consumer demand unfulfilled. Furthermore, existing techniques for marketing, advertising, or otherwise calling consumers to action failed to effectively utilize voice-based information, which is one of the most natural, intuitive methods of human interaction.

80. In certain embodiments, the '758 patent advantageously uses a speech recognition engine and natural language processing to interpret natural language utterances, determine context for the natural language utterance, select a purchase opportunity based on the determined context, deliver a selected purchase opportunity, tracking an interaction pattern associated with

the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity, wherein the subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity; building or updating a user-specific profile based on the tracked interaction pattern, and interpreting a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance. In doing so, the inventors were able to improve the functioning of voice user interface systems thereby improving the operation of those systems in an unconventional manner. For example, the innovations in the '758 patent allowed a user to directly issue natural language requests and engage in a productive, cooperative dialogue to resolve requests and advance a conversation. In this regard, one of the problems faced by the inventors was necessarily rooted in voice user interface technology specifically arising in the realm of voice user interface systems.

81. With respect to the '758 patent, the patent describes and claims a system for processing natural language utterances that include requests and selecting purchase opportunities. The system uses a speech recognition engine and a conversational language processor to interpret a natural language utterance, determine context, select and deliver a purchase opportunity, track an interaction pattern for the delivered purchase opportunity, build a user-specific profile based on the tracked interaction pattern, and interpreting a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity. Figure 2 shows an architecture for an exemplary advertising system:

82. Conventional wisdom lacked an appreciation for the substantial advantages associated with utilizing a speech recognition engine configured in this way with a conversational language processor to interpret a natural language utterance, determine context, select and deliver a purchase opportunity, track an interaction pattern for the delivered purchase opportunity, build a user-specific profile based on the tracked interaction pattern, and interpreting a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity. In this regard, using this speech recognition engine in a system that utilizes a conversational language processor in this way was not well-understood, routine, or conventional and stands in sharp contrast to the conventional and routine approach of existing systems that required a user to memorize specific syntaxes, words, phrases, concepts, semantic indicators, or other keywords/qualifiers.

83. During prosecution of the '758 patent, the examiner rejected numerous application claims as being unpatentable over Levin (U.S. Patent Application Pub. No. 2003/0061039). In response, the inventors' prosecution counsel argued that Levin fails to teach or suggest (i) "determining, by one or more processors, a context for the natural language utterance based on the recognized words or phrases," and (ii) "selecting, by the one or more processors, a purchase opportunity based on the determined context." The inventors' patent counsel provided the following example to help explain:

For example, in the *Levin* system, if a user wanted to search for a Mike Tyson pay-per-view fight, he might say, in response to successive queries from the system, “show titles,” then “Tyson fight,” and then “yes” to confirm that he wishes to order the fight. As can be understood from this example, each operation by the system is in direct response to a voice command from the user. There is no teaching or suggestion that this process determines a separate “context” outside of the stream of voice commands, or provides any response based on such a “context,” let alone the claimed “purchase opportunity.” Indeed, no context needs to be determined in the cited portions of *Levin* – the search results are all direct responses to voice commands. Further, there is no teaching or suggestion of the provision of a “purchase opportunity” in the cited portions of *Levin*. The only particular information provided to the user is the show that matches

While Applicant agrees that “context for the natural language utterance” does have a broad scope, it is not as broad as the Examiner asserts here. For example, as is apparent by reviewing claim 1 (and claim 11) a “context for the natural language utterance” is conceptually different from the “recognized words or phrases” of the natural language utterance – because it is information determined **based on** the words or phrases of the utterance. A good example of “context” information can be found in pars. [0032] and [0033] of the instant Application, where contexts (e.g., “navigation,” “sports,” “Philadelphia”) can be determined based on utterances, in addition to recognizing the underlying words or phrases. While *Levin* discloses recognizing words or phrases to complete its voice commands, it does not teach or suggest determining context from those words or phrases.

84. Based, at least in part, on the inventors’ arguments, the USPTO eventually granted the claims of the ’758 patent finding them novel and non-obvious over *Levin*. As indicated, a key distinction—which the examiner eventually agreed with—is that *Levin* lacked the use of “context” to advantageously enable the ability for “selecting...a purchase opportunity based on the determined context.”

85. Conventional wisdom in shopping was keenly focused on menu-based systems. Indeed, web-stores were (and still are) based on product categorizations. The typical user experience involves going through a series of menus to narrow down the particular product. As such, there was particular focus and motivation to emulate the menus in a speech-based system. That is, for the computer to read options which the end-user selects before moving to the next menu. The '758 patent represents a dramatic departure by “selecting, by the one or more processors, a purchase opportunity based on the determined context.”

86. A skilled artisan would not consider the claim limitations of the '758 patent – either alone or in combination – to recite well-understood, routine, or conventional concepts. Instead, a person of ordinary skill in the art would recognize that the claim limitations of the '758 patent are directed to the inventive concepts described in the specification and prosecution history.

The '025 Patent

87. United States Patent Number 9,502,025 (“the '025 patent”), entitled “System and Method for Providing a Natural Language Content Dedication Service,” was duly and legally issued on November 22, 2016, and names Mike Kennewick and Lynn Elise Armstrong as the inventors. Attached as Exhibit E is a true and correct copy of the '025 patent.

88. The '025 patent claims, among other things, a system for providing a natural language content dedication service, comprising: one or more processors; and one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to: receive a first utterance that includes a natural language utterance; determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance; provide the one or

more words or phrases as an input to a conversational language processor; interpret the first utterance, at the conversational language processor, based on the one or more words or phrases; identify, based on the interpretation of the first utterance, content to dedicate to a recipient; initiate, based on the identified content, a dedication to the recipient; receive a second utterance to be associated with the dedication; determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance; provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and send information to enable the recipient to access the content and the second utterance.

89. VB Assets, LLC is the assignee of the entire right, title, and interest in the '025 patent.

90. Voice user interface systems in existence before the invention of the '025 patent were typically of the “Command and Control” type. Such systems used verbal menus to restrict information that a person can provide at a given point. For example, the voice system would present the list of available options either verbally and/or on a screen. The user could then respond by speaking the menu item. Such a system could include numerous menus that the user would have to get through in order to convey the desired information to the system and/or to cause the system to take the desired action. The inventors recognized a significant problem with the Command and Control systems that caused 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 did 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.

91. Additionally, existing voice user interface systems were constrained to a finite set of applications or devices, limiting users' ability to access a wide array of applications and services across different devices as needed. Users often had to carry multiple devices to fulfill various needs, but the voice user interface systems did not effectively facilitate access to services and content across these disparate devices. There was a notable absence of an integrated environment that enabled users to request content or services associated with virtually any device or network, thereby restricting the availability of information and device interaction mechanisms. When users needed to perform a function on a device but were unsure how to do so, they could not simply use natural language to make the request. This resulted in simple functions becoming cumbersome and tedious. This could be illustrated by the example of purchasing a new ringtone for a mobile phone, which involved navigating through several menus and pressing numerous buttons. Additionally, the existing voice user interface systems did not support cooperative multi-modal interactions, which would allow users to engage with their devices in an intuitive, natural, and efficient manner. By addressing these constraints, the invention aims to enhance the user experience through advanced voice recognition and natural language processing capabilities, enabling more seamless and integrated interactions with electronic devices.

92. To overcome the shortcoming of prior art systems, the inventors provided a solution that provided a system and method for providing a natural language content dedication service operating in a voice services environment that can receive multi-modal natural language device interactions. In particular, providing the natural language content dedication service may

generally include detecting multi-modal device interactions that include requests to dedicate content, identifying the content requested for dedication from natural language utterances included in the multi-modal device interactions, processing transactions for the content requested for dedication, processing natural language to customize the content for recipients of the dedications, and delivering the customized content to the recipients of the dedications.

93. The advantages of the invention include providing a natural language content dedication service that operates in a voice services environment with electronic devices capable of receiving multi-modal natural language interactions. This system can detect interactions that include content dedication requests, identify the requested content from natural language utterances, process transactions, and customize the content for recipients. The invention supports a hybrid processing environment with multiple devices that cooperatively interpret and process natural language inputs, ensuring accurate intent determination. For example, a virtual router selects the cleanest audio sample of a natural language utterance and coordinates with other devices to resolve the request based on intent. This approach enhances user experience by simplifying interactions, making the technology intuitive and efficient. Users can engage in natural language dialogues to dedicate content, process transactions, and customize the content, all seamlessly integrated across various devices and domains. Additionally, the system supports flexible purchase options like buy-to-own, pay-to-play, and subscription models, allowing users to choose how they want to dedicate content, which can be customized and delivered efficiently.

94. During prosecution of the '025 patent, the examiner rejected numerous application claims as being unpatentable over prior art including Gold (U.S. Publication No. 2002/0032752) and Kennewick (U.S. Publication No. 2004/0193420). The USPTO eventually

granted the claims of the '025 patent finding them novel and non-obvious over the prior art, explaining:

However, the combination of prior art does not provide for the two step process of natural language processing during a dedication as reflected in the claim language of determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance; provide the one or more words or phrases as an input to a conversational language processor; interpret the first utterance, at the conversational language processor, based on the one or more words or phrases; determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance; and provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content."

95. A skilled artisan would not consider the claim limitations of the '025 patent – either alone or in combination – to recite well-understood, routine, or conventional concepts.

COUNT ONE (INFRINGEMENT OF '249 PATENT)

96. VoiceBox incorporates and realleges each and every allegation contained in the foregoing paragraphs of this complaint, as though fully set forth herein.

97. Amazon has and continues to infringe, literally and/or under the doctrine of equivalents, at least claims 1-10 and 12-15 of the '249 patent by making, using, selling, offering for sale, and/or importing into the United States, Alexa Products, which embody or use the inventions of the '249 patent in violation of 35 U.S.C. § 271(a). Exemplary evidence and an exemplary chart mapping a claim to Alexa Products can be found in Exhibit F.

98. Amazon has been and is inducing infringement of the '249 patent by actively and knowingly inducing others to make, use, sell, offer for sale, or import Alexa Products that include Alexa and embody or use the inventions claimed in the '249 patent in violation of 35 U.S.C. § 271(b). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, VoiceBox informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox's office of the '249 patent.³ On information and belief, Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that notified Amazon of its infringement of the '249 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

³ At the time, the '249 patent was referred to by its publication number: US 2015/0228276.

99. Amazon has been and is continuing to contributorily infringe the '249 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '249 patent and not a staple article or commodity of commerce suitable for substantial non-infringing use in violation of 35 U.S.C. § 271(c). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. Additionally, VoiceBox informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox's office of the '249 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that notified Amazon of its infringement of the '249 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than May of 2024. On information and belief, no later

than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use.

100. On information and belief, Amazon has known of the existence of the '249 patent, and its acts of infringement have been willful and in disregard for the '249 patent, without any reasonable basis for believing that it had a right to engage in the infringing conduct. Additionally, VoiceBox informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox's office of the '249 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '249 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '249 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '249 patent.

101. Amazon's infringement has been, and continues to be knowing, intentional, and willful.

102. Amazon's acts of infringement of the '249 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

103. This case is exceptional and, therefore, VoiceBox is entitled to an award of attorney's fees pursuant to 35 U.S.C. § 285.

COUNT TWO (INFRINGEMENT OF THE '699 PATENT)

104. VoiceBox incorporates and realleges each and every allegation contained in the foregoing paragraphs of this complaint, as though fully set forth herein.

105. Amazon has and continues to infringe, literally and/or under the doctrine of equivalents, at least claims 1, 3-5, 6-8, and 11 the '699 patent by making, using, selling, offering for sale, and/or importing into the United States, Alexa Products, which embody or use the inventions of the '699 patent in violation of 35 U.S.C. § 271(a). Exemplary evidence and an exemplary chart mapping a claim to Alexa Products can be found in Exhibit G.

106. Amazon has been and is inducing infringement of the '699 patent by actively and knowingly inducing others to make, use, sell, offer for sale, or import Alexa Products that include Alexa and embody or use the inventions claimed in the '699 patent in violation of 35 U.S.C. § 271(b). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '699 patent since no later than its date of issuance. No later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '699 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '699 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its actions would

induce infringement by users of Alexa Products. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

107. Amazon has been and is continuing to contributorily infringe the '699 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '699 patent and not a staple article or commodity of commerce suitable for substantial non-infringing use, in violation of 35 U.S.C. § 271(c). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '699 patent since no later than it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '699 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '699 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use.

108. On information and belief, Amazon has known of the existence of the '699 patent, and its acts of infringement have been willful and in disregard for the '699 patent, without any reasonable basis for believing that it had a right to engage in the infringing conduct. Amazon has known of, or been willfully blind to, the existence of the '699 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products infringe the '699 patent. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '699 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '699 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '699 patent.

109. Amazon's infringement has been, and continues to be knowing, intentional, and willful.

110. Amazon's acts of infringement of the '699 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

111. This case is exceptional and, therefore, VoiceBox is entitled to an award of attorney's fees pursuant to 35 U.S.C. § 285.

COUNT THREE (INFRINGEMENT OF THE '385 PATENT)

112. VoiceBox incorporates and realleges each and every allegation contained in the foregoing paragraphs of this complaint, as though fully set forth herein.

113. Amazon has and continues to infringe, literally and/or under the doctrine of equivalents, at least claims 1-5, 11-15 and 31-34 of the '385 patent by making, using, selling, offering for sale, and/or importing into the United States, Alexa Products, which embody or use

the inventions of the '385 patent in violation of 35 U.S.C. § 271(a). Exemplary evidence and an exemplary chart mapping a claim to Alexa Products can be found in Exhibit H.

114. Amazon has been and is inducing infringement of the '385 patent by actively and knowingly inducing others to make, use, sell, offer for sale, or import Alexa Products that include Alexa and embody or use the inventions claimed in the '385 patent, in violation of 35 U.S.C. § 271(b). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '385 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '385 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '385 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

115. Amazon has been and is continuing to contributorily infringe the '385 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '385 patent and not a staple article or commodity of commerce suitable for substantial non-infringing use in violation of 35 U.S.C. § 271(c). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products

to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '385 patent since no later than it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '385 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '385 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use.

116. On information and belief, Amazon has known of the existence of the '385 patent, and its acts of infringement have been willful and in disregard for the '385 patent, without any reasonable basis for believing that it had a right to engage in the infringing conduct. Amazon has known of, or been willfully blind to, the existence of the '385 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products infringe the '385 patent. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '385 patent. On

information and belief, Amazon has known of, or been willfully blind to, the existence of the '385 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '385 patent.

117. Amazon's infringement has been, and continues to be knowing, intentional, and willful.

118. Amazon's acts of infringement of the '385 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

119. This case is exceptional and, therefore, VoiceBox is entitled to an award of attorney's fees pursuant to 35 U.S.C. § 285.

COUNT FOUR (INFRINGEMENT OF THE '758 PATENT)

120. VoiceBox incorporates and realleges each and every allegation contained in the foregoing paragraphs of this complaint, as though fully set forth herein.

121. Amazon has and continues to infringe, literally and/or under the doctrine of equivalents, at least claims 1-7, 9-12, and 18-22 of the '758 patent by making, using, selling, offering for sale, and/or importing into the United States, Alexa Products, which embody or use the inventions of the '758 patent in violation of 35 U.S.C. § 271(a). Exemplary evidence and an exemplary chart mapping a claim to Alexa Products can be found in Exhibit I.

122. Amazon has been and is inducing infringement of the '758 patent by actively and knowingly inducing others to make, use, sell, offer for sale, or import Alexa Products that include Alexa and embody or use the inventions claimed in the '758 patent, in violation of 35 U.S.C. § 271(b). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be

made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '758 patent since its issuance. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '758 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '758 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

123. Amazon has been and is continuing to contributorily infringe the '758 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '758 patent and not a staple article or commodity of commerce suitable for substantial non-infringing use in violation of 35 U.S.C. § 271(c). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '758 patent since no later than it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both

patented and infringing and that there is no substantial non-infringing use. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '758 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '758 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use.

124. On information and belief, Amazon has known of the existence of the '758 patent, and its acts of infringement have been willful and in disregard for the '758 patent, without any reasonable basis for believing that it had a right to engage in the infringing conduct. Amazon has known of, or been willfully blind to, the existence of the '758 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products infringe the '758 patent. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '758 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '758 patent since no later than May of 2024. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '758 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '758 patent since no later

than that date. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '758 patent.

125. Amazon's infringement has been, and continues to be knowing, intentional, and willful.

126. Amazon's acts of infringement of the '758 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

127. This case is exceptional and, therefore, VoiceBox is entitled to an award of attorney's fees pursuant to 35 U.S.C. § 285.

COUNT FIVE (INFRINGEMENT OF THE '025 PATENT)

128. VoiceBox incorporates and realleges each and every allegation contained in the foregoing paragraphs of this complaint, as though fully set forth herein.

129. Amazon has and continues to infringe, literally and/or under the doctrine of equivalents, at least claims 8 and 14-19 the '025 patent by making, using, selling, offering for sale, and/or importing into the United States, Alexa Products, which embody or use the inventions of the '025 patent in violation of 35 U.S.C. § 271(a). Exemplary evidence and an exemplary chart mapping a claim to Alexa Products can be found in Exhibit J.

130. Amazon has been and is inducing infringement of the '025 patent by actively and knowingly inducing others to make, use, sell, offer for sale, or import Alexa Products that include Alexa and embody or use the inventions claimed in the '025 patent, in violation of 35 U.S.C. § 271(b). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from

third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '025 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than May of 2024. Additionally, the '025 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 9,953,630. The '025 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 4, 2018. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than January 4, 2018. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

131. Amazon has been and is continuing to contributorily infringe the '025 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '025 patent and not a staple article or commodity of commerce suitable for substantial non-infringing use, in violation of 35 U.S.C. § 271(c). On information and belief, Amazon writes software for Alexa Products and designs Alexa Products to operate in an infringing manner. Amazon causes Alexa Products to be made available through its own website. On information and belief, Amazon also profits from third-parties who sell Alexa Products. Amazon instructs users to use Alexa Products in an infringing manner and

provides technical support for such use, including on its website and also through the Alexa virtual assistant. Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '025 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than May of 2024. Additionally, the '025 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 9,953,630. The '025 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 4, 2018. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than January 4, 2018. On information and belief, no later than that date, Amazon has known its Alexa Products are especially made or adapted for a use or product that is both patented and infringing and that there is no substantial non-infringing use.

132. On information and belief, Amazon has known of the existence of the '025 patent, and its acts of infringement have been willful and in disregard for the '025 patent, without any reasonable basis for believing that it had a right to engage in the infringing conduct. Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than the date it received a copy of this complaint. No later than that date, Amazon has known its Alexa Products infringe the '025 patent. Additionally, in May 2024, VoiceBox sent a written summary of VoiceBox's patent portfolio to Amazon that informed Amazon of the '025 patent. On

information and belief, Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than May of 2024. Additionally, the '025 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 9,953,630. The '025 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 4, 2018. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '025 patent since no later than January 4, 2018. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '025 patent.

133. Amazon's infringement has been, and continues to be knowing, intentional, and willful.

134. Amazon's acts of infringement of the '025 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

135. This case is exceptional and, therefore, VoiceBox is entitled to an award of attorney's fees pursuant to 35 U.S.C. § 285.

PRAYER FOR RELIEF

WHEREFORE, VoiceBox, respectfully requests the Court to enter judgment in favor of VoiceBox and against Amazon as to all claims asserted herein as follows:

- a) Adjudging that Amazon has infringed, literally and/or under the doctrine of equivalents, actively induced infringement of, and contributorily infringed at least one claim of each of the VoiceBox Patents in violation of 35 U.S.C. §§ 271(a), (b), and/or (c);

- b) Ordering Amazon to account and pay damages adequate to compensate VoiceBox for Amazon's infringement of the VoiceBox Patents, including for any infringing acts not presented at trial and pre-judgment and post-judgment interest and costs, pursuant to 35 U.S.C. § 284;
- c) Ordering that the damages award be increased up to three times the actual amount assessed, pursuant to 35 U.S.C. § 284;
- d) Declaring this case exceptional and awarding VoiceBox its reasonable attorneys' fees, pursuant to 35 U.S.C. § 285;
- e) Awarding VoiceBox interest, including prejudgment and post-judgment interest, on the foregoing sums;
- f) Awarding VoiceBox the costs of this action; and
- g) Awarding such other and further relief as this Court deems just and proper.

JURY DEMAND

Pursuant to Federal Rules of Civil Procedure Rule 38, VoiceBox demands a trial by jury on all issues so triable.

Dated: July 18, 2024

Respectfully submitted,

SMITH, KATZENSTEIN & JENKINS LLP

Of counsel:

/s/ Neal C. Belgam

James C. Yoon
Ryan R. Smith
WILSON SONSINI GOODRICH
& ROSATI
650 Page Mill Road
Palo Alto, CA 94304
jyoon@wsgr.com
rsmith@wsgr.com

Neal C. Belgam (No. 2721)
Daniel A. Taylor (No. 6934)
1000 North West Street, Suite 1501
Wilmington, Delaware 19801
(302) 652-8400
nbelgam@skjlaw.com
dtaylor@skjlaw.com

Attorneys for Plaintiff VB Assets, LLC

Matthew A. Macdonald
Jamie Otto
953 East Third Street, Suite 100
Los Angeles, CA 90013
matthew.macdonald@wsgr.com
jotto@wsgr.com

Mikaela E. Evans-Aziz
One Market Plaza
Spear Tower, Suite 3300
San Francisco, CA 94105
mevansaziz@wsgr.com

David Greenbaum
GREENBAUM LAW LLC
210 Allison Court
Englewood, NJ 07631
david@greenbaum.law

EXHIBIT A

(12) **United States Patent**
Baldwin et al.

(10) **Patent No.:** **US 10,297,249 B2**
(45) **Date of Patent:** **May 21, 2019**

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

(56) **References Cited**

U.S. PATENT DOCUMENTS

(71) Applicant: **VB Assets, LLC**, Bellevue, WA (US)

4,430,669 A 2/1984 Cheung 358/122
4,821,027 A 4/1989 Mallory et al. 340/521

(Continued)

(72) Inventors: **Larry Baldwin**, Maple Valley, WA
(US); **Tom Freeman**, Mercer Island,
WA (US); **Michael Tjalve**, Bellevue,
WA (US); **Blane Ebersold**, Seattle, WA
(US); **Chris Weider**, Everett, WA (US)

FOREIGN PATENT DOCUMENTS

CN 1433554 A 7/2003
CN 1860496 A 11/2006

(Continued)

(73) Assignee: **VB Assets, LLC**, Bellevue, WA (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 289 days.

OTHER PUBLICATIONS

Kirchhoff, Katrin, "Syllable-Level Desynchronisation of Phonetic
Features for Speech Recognition", Proceedings of the Fourth Inter-
national Conference on Spoken Language, 1996, ICSLP 96, vol. 4,
IEEE, 1996, 3 pages.

(Continued)

(21) Appl. No.: **14/691,445**

(22) Filed: **Apr. 20, 2015**

(65) **Prior Publication Data**

US 2015/0228276 A1 Aug. 13, 2015

Related U.S. Application Data

(60) Continuation of application No. 13/987,645, filed on
Aug. 19, 2013, now Pat. No. 9,015,049, which is a
(Continued)

(51) **Int. Cl.**
G06F 3/16 (2006.01)
G06F 17/27 (2006.01)
(Continued)

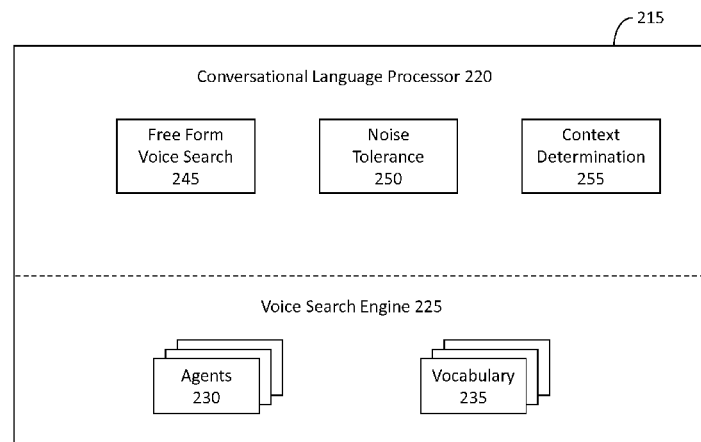
(52) **U.S. Cl.**
CPC **G10L 15/18** (2013.01); **G06F 3/167**
(2013.01); **G06F 17/2785** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC . G10L 15/22; G10L 15/1815; G10L 15/1822;
G10L 15/183; G10L 2015/225;
(Continued)

ABSTRACT

A cooperative conversational voice user interface is pro-
vided. 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 adap-
tive 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 imple-
mentation, misrecognitions may be tolerated, and conversa-
tional course may be corrected based on subsequent utter-
ances and/or responses.

28 Claims, 3 Drawing Sheets



US 10,297,249 B2

Page 2

Related U.S. Application Data						
	division of application No. 13/251,712, filed on Oct. 3, 2011, now Pat. No. 8,515,765, which is a continuation of application No. 11/580,926, filed on Oct. 16, 2006, now Pat. No. 8,073,681.		5,740,256 A	4/1998	Castello Da Costa et al.	361/94.7
			5,742,763 A	4/1998	Jones	395/200.3
			5,748,841 A *	5/1998	Morin	G06F 3/16
						704/257
			5,748,974 A	5/1998	Johnson	395/759
			5,752,052 A	5/1998	Richardson et al.	395/759
			5,754,784 A	5/1998	Garland et al.	395/200.49
			5,761,631 A	6/1998	Nasukawa	704/9
			5,774,841 A	6/1998	Salazar et al.	704/225
			5,774,859 A	6/1998	Houser et al.	704/275
			5,794,050 A	8/1998	Dahlgren et al.	395/708
			5,794,196 A	8/1998	Yegnanarayanan et al.	704/255
						701/201
			5,797,112 A	8/1998	Komatsu et al.	701/201
			5,799,276 A	8/1998	Komissarchik et al.	704/251
			5,802,510 A	9/1998	Jones	707/2
			5,829,000 A	10/1998	Huang	
			5,832,221 A	11/1998	Jones	375/200.36
			5,839,107 A	11/1998	Gupta et al.	704/270
			5,848,396 A	12/1998	Gerace	705/10
			5,855,000 A	12/1998	Waibel et al.	704/235
			5,860,059 A *	1/1999	Aust	G06F 3/16
						704/231
			5,867,817 A	2/1999	Catallo et al.	704/255
			5,878,385 A	3/1999	Bralich et al.	704/9
			5,878,386 A	3/1999	Coughlin	704/10
			5,892,813 A	4/1999	Morin et al.	379/88.01
			5,892,900 A	4/1999	Ginter et al.	395/186
			5,895,464 A	4/1999	Bhandari et al.	707/3
			5,895,466 A	4/1999	Goldberg et al.	707/5
			5,897,613 A	4/1999	Chan	704/210
			5,899,991 A *	5/1999	Karch	G06F 21/55
						707/694
			5,902,347 A	5/1999	Backman et al.	701/200
			5,911,120 A	6/1999	Jarett et al.	455/417
			5,918,222 A *	6/1999	Fukui	G06Q 10/10
			5,926,784 A	7/1999	Richardson et al.	704/9
			5,933,822 A	8/1999	Braden-Harder et al.	707/5
			5,950,167 A	9/1999	Yaker	704/275
			5,953,393 A	9/1999	Culbreth et al.	379/88.25
			5,960,384 A	9/1999	Brash	
			5,960,397 A	9/1999	Rahim	704/244
			5,960,399 A	9/1999	Barclay et al.	704/270
			5,960,447 A	9/1999	Holt et al.	707/500
			5,963,894 A	10/1999	Richardson et al.	704/9
			5,963,940 A	10/1999	Liddy et al.	707/5
			5,982,906 A *	11/1999	Ono	G10L 21/0208
						381/94.1
			5,983,190 A	11/1999	Trower, II	
			5,987,404 A	11/1999	Della Pietra et al.	704/9
			5,991,721 A	11/1999	Asano et al.	704/257
			5,995,119 A	11/1999	Cosatto et al.	345/473
			5,995,928 A	11/1999	Nguyen et al.	704/251
			5,995,943 A	11/1999	Bull	
			6,009,382 A	12/1999	Martino et al.	704/1
			6,014,559 A	1/2000	Amin	455/413
			6,018,708 A	1/2000	Dahan et al.	704/244
			6,021,384 A	2/2000	Gorin et al.	704/1
			6,028,514 A	2/2000	Lemelson et al.	340/539
			6,035,267 A	3/2000	Watanabe et al.	704/1
			6,044,347 A	3/2000	Abella et al.	704/272
			6,049,602 A	4/2000	Foladare et al.	379/265
			6,049,607 A	4/2000	Marash et al.	379/410
			6,058,187 A	5/2000	Chen	380/21
			6,067,513 A	5/2000	Ishimitsu	704/233
			6,073,098 A	6/2000	Buchsbaum	
			6,076,059 A	6/2000	Glickman et al.	704/260
			6,078,886 A *	6/2000	Dragosh	G10L 15/30
						704/243
			6,081,774 A	6/2000	De Hita et al.	704/9
			6,085,186 A	7/2000	Christianson et al.	707/3
			6,101,241 A	8/2000	Boyce et al.	379/88.01
			6,108,631 A	8/2000	Ruhl	704/270
			6,119,087 A	9/2000	Kuhn et al.	704/270
			6,119,101 A	9/2000	Peckover	
			6,122,613 A	9/2000	Baker	704/235
			6,134,235 A	10/2000	Goldman et al.	370/352
			6,144,667 A	11/2000	Doshi et al.	370/401
(51)	Int. Cl.					
	G10L 15/06 (2013.01)					
	G10L 15/18 (2013.01)					
	G10L 15/22 (2006.01)					
	G10L 17/22 (2013.01)					
	G10L 25/51 (2013.01)					
	G10L 15/183 (2013.01)					
	G10L 21/0216 (2013.01)					
(52)	U.S. Cl.					
	CPC G10L 15/1815 (2013.01); G10L 15/1822 (2013.01); G10L 15/22 (2013.01); G10L 17/22 (2013.01); G10L 25/51 (2013.01); G10L 15/183 (2013.01); G10L 2015/0631 (2013.01); G10L 2015/225 (2013.01); G10L 2015/228 (2013.01); G10L 2021/02166 (2013.01)					
(58)	Field of Classification Search					
	CPC G10L 2015/228; G10L 2021/02166; G10L 25/51; G06F 17/27; G06F 17/2785					
	See application file for complete search history.					
(56)	References Cited					
	U.S. PATENT DOCUMENTS					
	4,829,423 A 5/1989 Tennant et al.	364/200				
	4,887,212 A 12/1989 Zamora					
	4,910,784 A 3/1990 Doddington et al.	381/43				
	5,027,406 A 6/1991 Roberts et al.	381/43				
	5,155,743 A 10/1992 Jacobs	375/28				
	5,164,904 A 11/1992 Sumner	364/436				
	5,208,748 A 5/1993 Flores et al.	364/419				
	5,265,065 A 11/1993 Turtle					
	5,274,560 A 12/1993 LaRue	364/444				
	5,331,554 A 7/1994 Graham					
	5,357,596 A 10/1994 Takebayashi et al.	395/2.84				
	5,369,575 A 11/1994 Lamberti					
	5,377,350 A 12/1994 Skinner	395/600				
	5,386,556 A 1/1995 Hedin et al.	395/600				
	5,424,947 A 6/1995 Nagao et al.	364/419.08				
	5,471,318 A 11/1995 Ahuja et al.	358/400				
	5,475,733 A 12/1995 Eisdorfer et al.	379/52				
	5,479,563 A 12/1995 Yamaguchi					
	5,488,652 A 1/1996 Bielby et al.	379/88				
	5,499,289 A 3/1996 Bruno et al.	379/220				
	5,500,920 A 3/1996 Kupiec	395/2.79				
	5,517,560 A 5/1996 Greenspan	379/114				
	5,533,108 A 7/1996 Harris et al.	379/201				
	5,537,436 A 7/1996 Bottoms et al.	375/222				
	5,539,744 A 7/1996 Chu et al.	370/60				
	5,557,667 A 9/1996 Bruno et al.	379/201				
	5,559,864 A 9/1996 Kennedy, Jr.	379/59				
	5,563,937 A 10/1996 Bruno et al.	379/201				
	5,577,165 A 11/1996 Takebayashi et al.	395/2.84				
	5,590,039 A 12/1996 Ikeda et al.	395/759				
	5,608,635 A 3/1997 Tamai	364/449.3				
	5,615,296 A * 3/1997 Stanford	G10L 15/22				
		704/222				
	5,617,407 A 4/1997 Bareis	369/275.3				
	5,633,922 A 5/1997 August et al.	379/220				
	5,634,086 A 5/1997 Rtischev et al.	395/2.79				
	5,652,570 A 7/1997 Lepkofker	340/573				
	5,675,629 A 10/1997 Raffel et al.	379/58				
	5,696,965 A 12/1997 Dedrick	395/610				
	5,708,422 A 1/1998 Blonder et al.	340/825.34				
	5,721,938 A 2/1998 Stuckey	395/754				
	5,722,084 A 2/1998 Chakrin et al.	455/551				

US 10,297,249 B2

Page 3

(56)

References Cited

U.S. PATENT DOCUMENTS

6,144,938 A	11/2000	Surace et al.	704/257	6,434,523 B1	8/2002	Monaco	704/257
6,154,526 A	11/2000	Dahlke et al.	379/88.03	6,434,524 B1	8/2002	Weber	704/257
6,160,883 A	12/2000	Jackson et al.	379/230	6,434,529 B1	8/2002	Walker et al.	704/275
6,167,377 A	12/2000	Gillick et al.	704/240	6,442,522 B1	8/2002	Carberry et al.	704/257
6,173,266 B1	1/2001	Marx et al.	704/270	6,446,114 B1	9/2002	Bulfer et al.	709/206
6,173,279 B1	1/2001	Levin et al.	707/5	6,453,153 B1	9/2002	Bowker et al.	455/67.4
6,175,858 B1	1/2001	Bulfer et al.	709/206	6,453,292 B2	9/2002	Ramaswamy et al.	704/235
6,185,535 B1	2/2001	Hedin et al.	704/270	6,456,711 B1	9/2002	Cheung et al.	379/265.09
6,188,982 B1	2/2001	Chiang	704/256	6,456,974 B1	9/2002	Baker et al.	704/270.1
6,192,110 B1	2/2001	Abella et al.	379/88.01	6,466,654 B1	10/2002	Cooper et al.	379/88.01
6,192,338 B1	2/2001	Haszto et al.	704/257	6,466,899 B1	10/2002	Yano et al.	704/1
6,195,634 B1	2/2001	Dudemaine et al.	704/231	6,470,315 B1	10/2002	Netsch et al.	704/256
6,195,651 B1	2/2001	Handel et al.	707/2	6,487,494 B2	11/2002	Odinak et al.	701/202
6,199,043 B1	3/2001	Happ	704/272	6,487,495 B1	11/2002	Gale et al.	701/209
6,208,964 B1	3/2001	Sabourin	704/244	6,498,797 B1	12/2002	Anerousis et al.	370/522
6,208,972 B1	3/2001	Grant et al.	704/275	6,499,013 B1	12/2002	Weber	704/257
6,219,346 B1	4/2001	Maxemchuk	370/338	6,501,833 B2	12/2002	Phillips et al.	379/88.07
6,219,643 B1	4/2001	Cohen et al.	704/257	6,501,834 B1	12/2002	Milewski et al.	379/93.24
6,219,645 B1 *	4/2001	Byers	G10L 15/02 381/91	6,505,155 B1	1/2003	Vanbuskirk et al.	704/246
6,226,612 B1	5/2001	Srenger et al.	704/256	6,510,417 B1	1/2003	Woods et al.	704/275
6,233,556 B1	5/2001	Teunen et al.	704/250	6,513,006 B2	1/2003	Howard et al.	704/257
6,233,559 B1	5/2001	Balakrishnan	704/275	6,522,746 B1	2/2003	Marchok et al.	379/406.03
6,233,561 B1	5/2001	Junqua et al.	704/277	6,523,061 B1	2/2003	Halverson et al.	709/202
6,236,968 B1	5/2001	Kanevsky et al.	704/275	6,532,444 B1	3/2003	Weber	704/257
6,243,679 B1	6/2001	Mohri		6,539,348 B1	3/2003	Bond et al.	704/9
6,246,981 B1 *	6/2001	Papineni	G10L 15/22 704/235	6,549,629 B2	4/2003	Finn et al.	381/92
6,246,990 B1	6/2001	Happ	704/275	6,553,372 B1	4/2003	Brassell et al.	707/5
6,266,636 B1	7/2001	Kosaka et al.	704/244	6,556,970 B1 *	4/2003	Sasaki	G01C 21/3608 382/100
6,269,336 B1	7/2001	Ladd et al.	704/270	6,556,973 B1	4/2003	Lewin	704/277
6,272,455 B1	8/2001	Hoshen et al.	704/1	6,560,576 B1	5/2003	Cohen et al.	704/270
6,272,461 B1 *	8/2001	Meredith	G10L 15/26 704/235	6,560,590 B1	5/2003	Shwe et al.	706/55
6,275,231 B1	8/2001	Obradovich	345/349	6,567,778 B1	5/2003	Chao Chang et al.	704/257
6,278,377 B1	8/2001	DeLine et al.	340/815.4	6,567,797 B1	5/2003	Schuetze et al.	707/2
6,278,968 B1	8/2001	Franz et al.	704/3	6,567,805 B1	5/2003	Johnson	
6,286,002 B1	9/2001	Axaopoulos		6,570,555 B1	5/2003	Prevost et al.	345/156
6,288,319 B1	9/2001	Catona	84/609	6,570,964 B1	5/2003	Murveit et al.	379/67.1
6,292,767 B1	9/2001	Jackson et al.	704/1	6,571,279 B1	5/2003	Herz et al.	709/217
6,301,560 B1	10/2001	Masters	704/251	6,574,597 B1	6/2003	Mohri et al.	704/251
6,308,151 B1	10/2001	Smith	704/235	6,574,624 B1	6/2003	Johnson et al.	707/5
6,311,159 B1	10/2001	Van Tichelen et al.	704/275	6,578,022 B1	6/2003	Foulger et al.	706/45
6,314,402 B1	11/2001	Monaco et al.	704/275	6,581,103 B1	6/2003	Dengler	709/231
6,321,196 B1	11/2001	Franceschi	704/243	6,584,439 B1	6/2003	Geilhufe et al.	704/270
6,356,869 B1	3/2002	Chapados et al.	704/275	6,587,858 B1	7/2003	Strazza	707/102
6,362,748 B1	3/2002	Huang	340/901	6,591,185 B1	7/2003	Polidi	
6,366,882 B1	4/2002	Bijl et al.	704/235	6,591,239 B1	7/2003	McCall et al.	704/275
6,366,886 B1	4/2002	Dragosh et al.	704/270.1	6,594,257 B1	7/2003	Doshi et al.	370/352
6,374,214 B1	4/2002	Friedland et al.	704/235	6,594,367 B1	7/2003	Marash et al.	381/92
6,374,226 B1 *	4/2002	Hunt	G10L 15/26 704/255	6,598,018 B1	7/2003	Junqua	704/251
6,377,913 B1	4/2002	Coffman et al.	704/8	6,601,026 B2	7/2003	Appelt et al.	704/9
6,377,919 B1 *	4/2002	Burnett	A61B 5/0507 704/231	6,601,029 B1 *	7/2003	Pickering	G10L 15/22 704/257
6,381,535 B1	4/2002	Durocher et al.	701/202	6,604,075 B1	8/2003	Brown et al.	704/270.1
6,385,596 B1	5/2002	Wiser et al.	705/51	6,604,077 B2	8/2003	Dragosh et al.	704/270.1
6,385,646 B1	5/2002	Brown et al.	709/217	6,606,598 B1	8/2003	Holthouse et al.	704/275
6,389,398 B1 *	5/2002	Lustgarten	G10L 15/22 379/88.01	6,611,692 B2	8/2003	Raffel et al.	455/552
6,393,403 B1	5/2002	Majaniemi	704/275	6,614,773 B1	9/2003	Maxemchuk	370/337
6,393,428 B1	5/2002	Miller et al.	707/102	6,615,172 B1	9/2003	Bennett et al.	704/257
6,397,181 B1	5/2002	Li et al.	704/256	6,622,119 B1	9/2003	Ramaswamy et al.	704/9
6,404,878 B1	6/2002	Jackson et al.	379/221.01	6,629,066 B1	9/2003	Jackson et al.	704/9
6,405,170 B1	6/2002	Phillips et al.	704/270	6,631,346 B1	10/2003	Karaorman et al.	704/9
6,408,272 B1	6/2002	White et al.	704/270.1	6,631,351 B1	10/2003	Ramachandran et al.	704/270
6,411,810 B1	6/2002	Maxemchuk	455/453	6,633,846 B1	10/2003	Bennett et al.	704/257
6,411,893 B2	6/2002	Ruhl	701/207	6,636,790 B1	10/2003	Lightner et al.	701/33
6,415,257 B1	7/2002	Junqua et al.	704/275	6,643,620 B1	11/2003	Contolini et al.	704/270
6,418,210 B1	7/2002	Sayko	379/142.15	6,647,363 B2	11/2003	Claassen	704/1
6,420,975 B1	7/2002	DeLine et al.	340/815.4	6,650,747 B1	11/2003	Bala et al.	379/265.06
6,429,813 B2	8/2002	Feigen	342/357.13	6,658,388 B1 *	12/2003	Kleindienst	G10L 13/00 704/270
6,430,285 B1	8/2002	Bauer et al.	379/265.01	6,678,680 B1	1/2004	Woo	707/6
6,430,531 B1	8/2002	Polish	704/257	6,681,206 B1	1/2004	Gorin et al.	704/243
				6,691,151 B1	2/2004	Cheyser et al.	709/202
				6,701,294 B1	3/2004	Ball et al.	704/257
				6,704,396 B2	3/2004	Parolkar et al.	379/88.17
				6,704,576 B1	3/2004	Brachman et al.	455/503
				6,704,708 B1	3/2004	Pickering	704/235
				6,707,421 B1	3/2004	Drury et al.	342/357.1
				6,708,150 B1	3/2004	Hirayama et al.	704/243
				6,721,001 B1	4/2004	Berstis	348/231.3

US 10,297,249 B2

Page 4

(56)

References Cited

U.S. PATENT DOCUMENTS

6,721,633 B2	4/2004	Funk et al.	701/1	7,024,364 B2	4/2006	Guerra et al.	704/270
6,721,706 B1 *	4/2004	Strubbe	G10L 15/18	7,027,586 B2	4/2006	Bushey et al.	379/265.09
			704/257	7,027,974 B1	4/2006	Busch	
6,726,636 B2	4/2004	Der Ghazarian et al.	600/532	7,027,975 B1	4/2006	Pazandak et al.	704/9
6,732,088 B1 *	5/2004	Glance	G06F 17/30864	7,031,908 B1	4/2006	Huang	
6,735,592 B1	5/2004	Neumann et al.	707/101	7,035,415 B2	4/2006	Belt et al.	381/92
6,739,556 B1	5/2004	Langston	244/189	7,036,128 B1	4/2006	Julia et al.	719/317
6,741,931 B1	5/2004	Kohut et al.	701/209	7,043,425 B2	5/2006	Pao	704/211
6,742,021 B1	5/2004	Halverson et al.	709/218	7,054,817 B2	5/2006	Shao	704/270
6,745,161 B1	6/2004	Arnold et al.	704/7	7,058,890 B2	6/2006	George et al.	715/728
6,751,591 B1 *	6/2004	Gorin	G10L 15/22	7,062,488 B1	6/2006	Reisman	707/8
			379/88.01	7,069,220 B2	6/2006	Coffman et al.	704/275
6,751,612 B1	6/2004	Schuetze et al.	707/4	7,072,834 B2	7/2006	Zhou	704/244
6,754,485 B1	6/2004	Obradovich et al.	455/414.1	7,072,888 B1 *	7/2006	Perkins	G06F 17/30867
6,754,627 B2	6/2004	Woodward	704/235				
6,754,647 B1 *	6/2004	Tackett	G06F 17/279	7,076,362 B2	7/2006	Ohtsuji et al.	707/733
			700/246	7,082,469 B2	7/2006	Gold et al.	709/231
6,757,544 B2	6/2004	Rangarajan et al.	455/456.1	7,085,708 B2	8/2006	Manson	704/9
6,757,718 B1	6/2004	Halverson et al.	709/218	7,092,928 B1	8/2006	Elad et al.	706/60
6,785,651 B1 *	8/2004	Wang	G06F 9/4446	7,107,210 B2	9/2006	Deng et al.	704/226
			379/88.01	7,107,218 B1	9/2006	Preston	704/270
6,795,808 B1	9/2004	Strubbe et al.	704/275	7,110,951 B1	9/2006	Lemelson et al.	704/270
6,801,604 B2	10/2004	Maes et al.	379/88.17	7,127,395 B1	10/2006	Gorin et al.	704/257
6,801,893 B1	10/2004	Backfried et al.	704/257	7,127,400 B2	10/2006	Koch	704/270.1
6,804,330 B1 *	10/2004	Jones	H04M 3/4938	7,130,390 B2	10/2006	Abhuri	379/88.17
			379/88.01	7,136,875 B2	11/2006	Anderson et al.	707/104.1
6,810,375 B1	10/2004	Ejerhed		7,137,126 B1 *	11/2006	Coffman	G06F 17/30899
6,813,341 B1	11/2004	Mahoney	379/88.01				704/200
6,816,830 B1	11/2004	Kempe		7,143,037 B1	11/2006	Chestnut	704/251
6,823,308 B2 *	11/2004	Keiller	G10L 15/32	7,143,039 B1	11/2006	Stifelman et al.	704/270
			704/236	7,146,315 B2 *	12/2006	Balan	G10L 25/78
6,829,603 B1	12/2004	Chai et al.	707/5				704/233
6,832,230 B1	12/2004	Zilliacus et al.	707/203	7,146,319 B2	12/2006	Hunt	704/254
6,833,848 B1	12/2004	Wolff et al.	345/719	7,149,696 B2	12/2006	Shimizu et al.	705/1
6,850,603 B1	2/2005	Eberle et al.	379/88.16	7,165,028 B2	1/2007	Gong	704/233
6,856,990 B2	2/2005	Barile et al.	707/10	7,170,993 B2	1/2007	Anderson et al.	379/265.09
6,865,481 B2	3/2005	Kawazoe et al.	701/211	7,171,291 B2	1/2007	Obradovich	701/29
6,868,380 B2	3/2005	Kroeker	704/240	7,174,300 B2	2/2007	Bush	704/275
6,868,385 B1	3/2005	Gerson	704/275	7,177,798 B2	2/2007	Hsu et al.	704/9
6,871,179 B1	3/2005	Kist		7,184,957 B2	2/2007	Brookes et al.	704/246
6,873,837 B1	3/2005	Yoshioka et al.	455/321	7,190,770 B2	3/2007	Ando et al.	379/88.01
6,877,001 B2	4/2005	Wolf et al.	707/3	7,197,069 B2	3/2007	Agazzi et al.	375/233
6,877,134 B1	4/2005	Fuller et al.	715/500.1	7,197,460 B1	3/2007	Gupta et al.	704/270.1
6,882,970 B1 *	4/2005	Garner	G06F 17/30746	7,203,644 B2	4/2007	Anderson et al.	704/246
			704/236	7,206,418 B2	4/2007	Yang et al.	381/92
6,901,366 B1	5/2005	Kuhn et al.	704/275	7,207,011 B2	4/2007	Mulvey et al.	715/812
6,910,003 B1	6/2005	Arnold et al.	704/4	7,215,941 B2	5/2007	Beckmann et al.	455/404.1
6,912,498 B2	6/2005	Stevens et al.	704/235	7,228,276 B2	6/2007	Omote et al.	704/243
6,915,126 B2	7/2005	Mazzara, Jr.	455/411	7,231,343 B1	6/2007	Treadgold et al.	704/9
6,928,614 B1	8/2005	Everhart	715/728	7,236,923 B1	6/2007	Gupta	704/9
6,934,756 B2	8/2005	Maes	709/227	7,254,482 B2	8/2007	Kawasaki et al.	701/211
6,937,977 B2	8/2005	Gerson	704/201	7,272,212 B2	9/2007	Eberle et al.	379/88.17
6,937,982 B2	8/2005	Kitaoka et al.	704/252	7,277,854 B2	10/2007	Bennett et al.	704/257
6,941,266 B1	9/2005	Gorin	704/257	7,283,829 B2	10/2007	Christenson et al.	455/461
6,944,594 B2	9/2005	Busayapongchai et al.		7,283,951 B2	10/2007	Marchisio et al.	704/9
			704/275	7,289,606 B2	10/2007	Sibal et al.	379/52
6,950,821 B2	9/2005	Faybishenko et al.	707/10	7,299,186 B2	11/2007	Kuzunuki et al.	704/270.1
6,954,755 B2	10/2005	Reisman	707/10	7,301,093 B2	11/2007	Sater et al.	84/615
6,959,276 B2	10/2005	Droppo et al.	704/226	7,305,381 B1	12/2007	Poppink et al.	
6,961,700 B2	11/2005	Mitchell et al.	704/235	7,321,850 B2	1/2008	Wakita	704/10
6,963,759 B1	11/2005	Gerson	455/563	7,328,155 B2	2/2008	Endo et al.	704/251
6,964,023 B2	11/2005	Maes et al.	715/811	7,337,116 B2	2/2008	Charlesworth et al.	704/254
6,968,311 B2	11/2005	Knockeart et al.	704/270	7,340,040 B1	3/2008	Saylor et al.	379/67.1
6,973,387 B2	12/2005	Masclet et al.	701/211	7,366,285 B2	4/2008	Parolkar et al.	379/88.17
6,975,983 B1 *	12/2005	Fortescue	G06F 3/038	7,366,669 B2	4/2008	Nishitani et al.	704/256
			704/257	7,376,645 B2	5/2008	Bernard	707/3
6,975,993 B1	12/2005	Keiller	704/275	7,380,250 B2	5/2008	Schechter	
6,980,092 B2	12/2005	Turnbull et al.	340/425.5	7,386,443 B1	6/2008	Parthasarathy et al.	704/201
6,983,055 B2	1/2006	Luo	381/313	7,398,209 B2	7/2008	Kennewick et al.	704/255
6,990,513 B2	1/2006	Belfiore et al.	709/203	7,406,421 B2	7/2008	Odinak et al.	704/275
6,996,531 B1	2/2006	Korall et al.	704/270	7,415,100 B2	8/2008	Cooper	
7,003,463 B2	2/2006	Maes et al.	704/270.1	7,415,414 B2	8/2008	Azara et al.	704/270
7,016,849 B2	3/2006	Arnold et al.	704/275	7,421,393 B1	9/2008	Di Fabbri et al.	704/275
7,020,609 B2	3/2006	Thrift et al.	704/270.1	7,424,431 B2	9/2008	Greene et al.	704/270
				7,447,635 B1	11/2008	Konopka et al.	704/275
				7,451,088 B1	11/2008	Ehlen et al.	704/270.1
				7,454,368 B2	11/2008	Stillman	
				7,454,608 B2	11/2008	Gopalakrishnan et al.	713/100
				7,461,059 B2	12/2008	Richardson et al.	707/5

US 10,297,249 B2

Page 5

(56)

References Cited

U.S. PATENT DOCUMENTS

7,472,020 B2	12/2008	Brulle-Drews	701/211	8,069,046 B2	11/2011	Kennewick et al.	704/257
7,472,060 B1	12/2008	Gorin et al.	704/240	8,073,681 B2	12/2011	Baldwin et al.	704/9
7,472,075 B2	12/2008	Odinak et al.	705/26	8,077,975 B2	12/2011	Ma et al.	382/187
7,477,909 B2	1/2009	Roth	455/466	8,082,153 B2	12/2011	Coffman et al.	704/270
7,478,036 B2	1/2009	Shen et al.	704/9	8,086,463 B2	12/2011	Ativanichayaphong et al.	704/275
7,487,088 B1	2/2009	Gorin et al.	704/240	8,103,510 B2	1/2012	Sato	
7,487,110 B2	2/2009	Bennett et al.	705/26	8,112,275 B2	2/2012	Kennewick et al.	704/240
7,493,259 B2	2/2009	Jones et al.	704/257	8,140,327 B2	3/2012	Kennewick et al.	704/226
7,493,559 B1	2/2009	Wolff et al.	715/727	8,140,335 B2	3/2012	Kennewick et al.	704/257
7,502,672 B1	3/2009	Kolls	701/29	8,145,489 B2	3/2012	Freeman et al.	704/257
7,502,730 B2	3/2009	Wang		8,150,694 B2	4/2012	Kennewick et al.	704/257
7,502,738 B2	3/2009	Kennewick et al.	704/257	8,155,962 B2	4/2012	Kennewick et al.	704/257
7,512,906 B1	3/2009	Baier		8,170,867 B2	5/2012	Germain	704/9
7,516,076 B2	4/2009	Walker et al.	704/275	8,180,037 B1	5/2012	Delker	
7,529,675 B2	5/2009	Maes	704/270.1	8,195,468 B2	6/2012	Weider et al.	704/275
7,536,297 B2	5/2009	Byrd et al.	704/10	8,200,485 B1	6/2012	Lee et al.	704/231
7,536,374 B2	5/2009	Au	706/55	8,204,751 B1 *	6/2012	Di Fabbriozio	G06F 17/271
7,542,894 B2	6/2009	Murata	704/9	8,219,399 B2	7/2012	Lutz et al.	704/270
7,546,382 B2	6/2009	Healey et al.	709/246	8,219,599 B2	7/2012	Tunstall-Pedoe	707/955
7,548,491 B2	6/2009	Macfarlane	367/198	8,224,652 B2	7/2012	Wang et al.	704/275
7,552,054 B1	6/2009	Stifelman et al.	704/270	8,255,224 B2	8/2012	Singleton et al.	704/275
7,558,730 B2	7/2009	Davis et al.	704/235	8,326,599 B2	12/2012	Tomeh	
7,574,362 B2	8/2009	Walker et al.	704/275	8,326,627 B2	12/2012	Kennewick et al.	704/257
7,577,244 B2	8/2009	Taschereau	379/218.01	8,326,634 B2	12/2012	Di Cristo et al.	704/270.1
7,606,708 B2	10/2009	Hwang	704/257	8,326,637 B2	12/2012	Baldwin et al.	704/275
7,606,712 B1 *	10/2009	Smith	G06F 3/167	8,332,224 B2	12/2012	Di Cristo et al.	704/257
7,620,549 B2	11/2009	Di Cristo et al.	704/257	8,340,975 B1	12/2012	Rosenberger	
7,634,409 B2	12/2009	Kennewick et al.	704/257	8,346,563 B1	1/2013	Hjelm	
7,640,006 B2	12/2009	Portman et al.	455/412.1	8,370,147 B2	2/2013	Kennewick et al.	704/257
7,640,160 B2	12/2009	Di Cristo et al.	704/257	8,447,607 B2	5/2013	Weider et al.	704/250
7,640,272 B2	12/2009	Mahajan et al.	707/104.1	8,447,651 B1	5/2013	Scholl	
7,672,931 B2	3/2010	Hurst-Hiller		8,452,598 B2	5/2013	Kennewick et al.	704/257
7,676,365 B2	3/2010	Hwang et al.	704/240	8,503,995 B2	8/2013	Ramer	
7,676,369 B2	3/2010	Fujimoto et al.	704/270	8,509,403 B2	8/2013	Chiu et al.	379/114.13
7,684,977 B2	3/2010	Morikawa	704/211	8,515,765 B2	8/2013	Baldwin et al.	704/275
7,693,720 B2	4/2010	Kennewick et al.	704/275	8,527,274 B2	9/2013	Freeman et al.	704/257
7,697,673 B2	4/2010	Chiu		8,577,671 B1	11/2013	Barve	
7,706,616 B2	4/2010	Kristensson		8,589,161 B2	11/2013	Kennewick et al.	704/252
7,729,913 B1 *	6/2010	Lee	G06F 17/30654	8,612,205 B2	12/2013	Hanneman	
7,729,916 B2	6/2010	Coffman et al.	704/270	8,612,206 B2	12/2013	Chalabi	
7,729,918 B2	6/2010	Walker et al.	704/275	8,620,659 B2	12/2013	Di Cristo et al.	704/257
7,729,920 B2	6/2010	Cheer et al.	704/275	8,719,005 B1	5/2014	Lee	
7,734,287 B2	6/2010	Ying	455/423	8,719,009 B2	5/2014	Baldwin et al.	704/9
7,748,021 B2	6/2010	Obradovich	725/105	8,719,026 B2	5/2014	Kennewick et al.	704/257
7,788,084 B2	8/2010	Brun et al.	704/7	8,731,929 B2	5/2014	Kennewick et al.	704/257
7,792,257 B1	9/2010	Vanier		8,738,380 B2	5/2014	Baldwin et al.	704/257
7,801,731 B2	9/2010	Odinak et al.	704/275	8,849,652 B2	9/2014	Weider et al.	704/9
7,809,570 B2	10/2010	Kennewick et al.	704/257	8,849,670 B2	9/2014	Di Cristo et al.	704/270.1
7,818,176 B2	10/2010	Freeman et al.	704/270	8,849,696 B2	9/2014	Pansari	
7,831,426 B2	11/2010	Bennett	704/252	8,849,791 B1	9/2014	Hertschuh	
7,831,433 B1	11/2010	Belvin et al.	704/275	8,886,536 B2	11/2014	Freeman et al.	704/257
7,856,358 B2	12/2010	Ho	704/270	8,972,243 B1	3/2015	Strom	
7,873,519 B2	1/2011	Bennett	704/257	8,983,839 B2	3/2015	Kennewick et al.	704/254
7,873,523 B2	1/2011	Potter et al.	704/275	9,009,046 B1	4/2015	Stewart	
7,873,654 B2	1/2011	Bernard	707/769	9,015,049 B2	4/2015	Baldwin	
7,881,936 B2	2/2011	Longe et al.	704/257	9,037,455 B1	5/2015	Faaborg	
7,890,324 B2	2/2011	Bangalore et al.	704/231	9,070,366 B1	6/2015	Mathias	
7,894,849 B2	2/2011	Kass et al.	455/550.1	9,070,367 B1 *	6/2015	Hoffmeister	G10L 15/187
7,902,969 B2	3/2011	Obradovich	340/439	9,105,266 B2	8/2015	Baldwin	
7,917,367 B2	3/2011	Di Cristo et al.	704/270.1	9,171,541 B2	10/2015	Kennewick	
7,920,682 B2	4/2011	Byrne et al.	379/88.18	9,269,097 B2	2/2016	Freeman	
7,949,529 B2	5/2011	Weider et al.	704/270	9,305,548 B2	4/2016	Kennewick	
7,949,537 B2	5/2011	Walker et al.	704/275	9,308,445 B1	4/2016	Merzenich	
7,953,732 B2	5/2011	Frank et al.	707/724	9,318,108 B2	4/2016	Gruber	
7,974,875 B1	7/2011	Quilici et al.	705/14.4	9,406,078 B2	8/2016	Freeman	
7,983,917 B2	7/2011	Kennewick et al.	704/257	9,443,514 B1 *	9/2016	Taubman	G10L 15/22
7,984,287 B2	7/2011	Gopalakrishnan et al.	713/100	9,502,025 B2	11/2016	Kennewick	
8,005,683 B2	8/2011	Tessel et al.	704/275	2001/0021905 A1 *	9/2001	Burnett	A61B 5/0507
8,015,006 B2	9/2011	Kennewick et al.	704/236				704/233
8,024,186 B1	9/2011	De Bonet		2001/0039492 A1	11/2001	Nemoto	704/231
8,027,965 B2	9/2011	Takehara		2001/0041980 A1	11/2001	Howard et al.	704/270
8,032,383 B1	10/2011	Bhardwaj		2001/0047261 A1 *	11/2001	Kassan	G10L 15/30
8,060,367 B2	11/2011	Keaveney	704/247				704/270
				2001/0049601 A1	12/2001	Kroeker et al.	704/254
				2001/0054087 A1	12/2001	Flom et al.	709/218
				2002/0002548 A1 *	1/2002	Roundtree	G06F 17/30014

US 10,297,249 B2

Page 6

(56)	References Cited			2003/0158731 A1	8/2003	Falcon et al.	704/231
	U.S. PATENT DOCUMENTS			2003/0161448 A1	8/2003	Parolkar et al.	379/88.17
				2003/0167167 A1 *	9/2003	Gong	G10L 15/22
							704/250
2002/0007267 A1 *	1/2002	Batchilo	G06F 17/30663	2003/0174155 A1	9/2003	Weng	
			704/9	2003/0182132 A1	9/2003	Niemoeller	704/275
2002/0010584 A1	1/2002	Schultz		2003/0187643 A1	10/2003	VanThong	
2002/0015500 A1	2/2002	Belt et al.	381/66	2003/0204492 A1	10/2003	Wolf et al.	707/3
2002/0022927 A1	2/2002	Lemelson et al.	701/301	2003/0206640 A1	11/2003	Malvar et al.	381/93
2002/0022956 A1	2/2002	Ukrainczyk		2003/0212550 A1	11/2003	Ubale	704/215
2002/0029186 A1	3/2002	Roth		2003/0212558 A1	11/2003	Matula	704/260
2002/0029261 A1	3/2002	Shibata	709/219	2003/0212562 A1	11/2003	Patel et al.	704/275
2002/0032752 A1	3/2002	Gold et al.	709/218	2003/0225825 A1	12/2003	Healey et al.	709/203
2002/0035501 A1	3/2002	Handel et al.	705/10	2003/0233230 A1 *	12/2003	Ammicht	G10L 15/183
2002/0040297 A1	4/2002	Tsiao et al.	704/257				704/235
2002/0049535 A1	4/2002	Rigo et al.	701/211	2003/0236664 A1	12/2003	Sharma	704/251
2002/0049805 A1	4/2002	Yamada et al.	709/202	2004/0006475 A1 *	1/2004	Ehlen	G10L 15/22
2002/0059068 A1	5/2002	Rose et al.	704/246				704/270.1
2002/0065568 A1	5/2002	Silfvast et al.	700/94	2004/0010358 A1	1/2004	Oesterling et al.	701/49
2002/0067839 A1	6/2002	Heinrich	381/110	2004/0025115 A1	2/2004	Sienel et al.	715/513
2002/0069059 A1	6/2002	Smith	704/257	2004/0030741 A1	2/2004	Wolton	
2002/0069071 A1	6/2002	Knockeart et al.	704/275	2004/0036601 A1	2/2004	Obradovich	
2002/0073176 A1	6/2002	Ikeda		2004/0044516 A1 *	3/2004	Kennewick	G10L 15/22
2002/0082911 A1	6/2002	Dunn et al.	705/14				704/5
2002/0087312 A1 *	7/2002	Lee	G06Q 30/06	2004/0093567 A1	5/2004	Schabes	
			704/251	2004/0098245 A1	5/2004	Walker et al.	704/1
2002/0087326 A1	7/2002	Lee et al.	704/270.1	2004/0102977 A1	5/2004	Metzler	
2002/0087525 A1	7/2002	Abbott et al.	707/3	2004/0117179 A1	6/2004	Balasuriya	704/231
2002/0107694 A1	8/2002	Lerg	704/273	2004/0117804 A1	6/2004	Scahill et al.	719/320
2002/0120609 A1	8/2002	Lang et al.	707/1	2004/0122673 A1 *	6/2004	Park	G10L 15/24
2002/0124050 A1	9/2002	Middeljans	709/203				704/270
2002/0133347 A1 *	9/2002	Schoneburg	G06F 17/27	2004/0122674 A1	6/2004	Bangalore et al.	704/276
			704/257	2004/0133793 A1	7/2004	Ginter	
2002/0133354 A1 *	9/2002	Ross	G10L 15/30	2004/0140989 A1	7/2004	Papageorge	345/700
			704/275	2004/0143440 A1	7/2004	Prasad	
2002/0133402 A1	9/2002	Faber et al.	705/14	2004/0148154 A1	7/2004	Acero	
2002/0135618 A1 *	9/2002	Maes	G06F 3/0481	2004/0148170 A1	7/2004	Acero	
			715/767	2004/0158555 A1	8/2004	Seedman et al.	707/3
2002/0138248 A1	9/2002	Corston-Oliver et al.	704/1	2004/0166832 A1	8/2004	Portman et al.	455/412.1
2002/0143532 A1	10/2002	McLean et al.	704/235	2004/0167771 A1	8/2004	Duan et al.	704/10
2002/0143535 A1	10/2002	Kist et al.	704/251	2004/0172247 A1	9/2004	Yoon	
2002/0152260 A1	10/2002	Chen		2004/0172258 A1	9/2004	Dominach et al.	704/277
2002/0161646 A1	10/2002	Gailey et al.	705/14	2004/0189697 A1	9/2004	Fukuoka	
2002/0161647 A1	10/2002	Gailey		2004/0193408 A1	9/2004	Hunt	704/209
2002/0169597 A1 *	11/2002	Fain	G10L 15/22	2004/0193420 A1 *	9/2004	Kennewick	G06F 17/30654
			704/9				704/257
2002/0173333 A1	11/2002	Buchholz et al.	455/527	2004/0199375 A1	10/2004	Ehsani et al.	704/4
2002/0173961 A1	11/2002	Guerra	704/258	2004/0199389 A1 *	10/2004	Geiger	G10L 15/16
2002/0178003 A1 *	11/2002	Gehrke	G10L 15/30				704/254
			704/246	2004/0201607 A1 *	10/2004	Mulvey	G06F 3/0237
2002/0184373 A1	12/2002	Maes	709/228				715/708
2002/0188455 A1 *	12/2002	Shioda	G10L 17/26	2004/0204939 A1 *	10/2004	Liu	G10L 15/28
			704/275				704/239
2002/0188602 A1	12/2002	Stubler et al.	707/3	2004/0205671 A1	10/2004	Sukehiro et al.	715/532
2002/0198714 A1	12/2002	Zhou	704/252	2004/0220800 A1 *	11/2004	Kong	G10L 21/0208
2003/0005033 A1 *	1/2003	Mohan	G06F 17/30899				704/205
			709/202	2004/0243393 A1 *	12/2004	Wang	G10L 15/22
2003/0014261 A1	1/2003	Kageyama	704/275				704/9
2003/0016835 A1	1/2003	Elko et al.	381/92	2004/0243417 A9	12/2004	Pitts, III et al.	704/276
2003/0036903 A1 *	2/2003	Konopka	G10L 15/065	2004/0247092 A1	12/2004	Timmins	
			704/249	2004/0249636 A1 *	12/2004	Applebaum	G10L 15/1822
2003/0046071 A1 *	3/2003	Wyman	G10L 15/22				704/231
			704/235	2005/0015256 A1	1/2005	Kargman	704/272
2003/0046281 A1 *	3/2003	Son	G06F 17/30876	2005/0021331 A1	1/2005	Huang et al.	704/231
2003/0046346 A1	3/2003	Mumick et al.	709/205	2005/0021334 A1	1/2005	Iwahashi	704/240
2003/0064709 A1	4/2003	Gailey et al.	455/412	2005/0021470 A1	1/2005	Martin et al.	705/51
2003/0065427 A1	4/2003	Funk et al.	701/1	2005/0021826 A1	1/2005	Kumar	709/232
2003/0069734 A1	4/2003	Everhart	704/275	2005/0033574 A1	2/2005	Kim et al.	704/251
2003/0069880 A1	4/2003	Harrison		2005/0033582 A1	2/2005	Gadd et al.	704/277
2003/0088421 A1	5/2003	Maes et al.	704/270.1	2005/0043940 A1	2/2005	Elder	704/9
2003/0093419 A1 *	5/2003	Bangalore	G01C 21/3664	2005/0060142 A1 *	3/2005	Visser	G10L 21/0208
2003/0097249 A1	5/2003	Walker et al.	704/1				704/201
2003/0110037 A1	6/2003	Walker et al.	704/257	2005/0080632 A1	4/2005	Endo et al.	704/277
2003/0112267 A1	6/2003	Belrose	345/728	2005/0102282 A1 *	5/2005	Linden	G06F 17/3053
2003/0115062 A1	6/2003	Walker et al.	704/258	2005/0114116 A1	5/2005	Fiedler	704/201
2003/0120493 A1	6/2003	Gupta	704/270.1	2005/0125232 A1	6/2005	Gadd	704/270.1
2003/0135488 A1	7/2003	Amir et al.	707/3	2005/0131673 A1	6/2005	Koizumi et al.	704/2
2003/0144846 A1	7/2003	Denenberg et al.	704/277	2005/0137850 A1	6/2005	Odell	704/4

US 10,297,249 B2

Page 7

(56)	References Cited			2007/0179778	A1	8/2007	Gong et al.	704/9
	U.S. PATENT DOCUMENTS			2007/0185859	A1	8/2007	Flowers	
				2007/0186165	A1	8/2007	Maislos et al.	715/728
				2007/0192309	A1	8/2007	Fischer	
2005/0137877	A1	6/2005	Oesterling et al.	2007/0198267	A1	8/2007	Jones et al.	704/257
2005/0143994	A1	6/2005	Mori et al.	2007/0203699	A1	8/2007	Nagashima	
2005/0144013	A1 *	6/2005	Fujimoto	2007/0203736	A1	8/2007	Ashton	705/1
				2007/0208732	A1	9/2007	Flowers	
2005/0144187	A1	6/2005	Che	2007/0214182	A1	9/2007	Rosenberg	707/104.1
2005/0149319	A1	7/2005	Honda	2007/0250901	A1	10/2007	McIntire et al.	725/146
2005/0169441	A1 *	8/2005	Yacoub	2007/0265850	A1	11/2007	Kennewick et al.	704/257
				2007/0266257	A1	11/2007	Camaisa	
				2007/0276651	A1	11/2007	Bliss	
2005/0216254	A1	9/2005	Gupta et al.	2007/0294615	A1 *	12/2007	Sathe	G06F 17/3070
2005/0222763	A1	10/2005	Uyeki					707/803
2005/0234637	A1	10/2005	Obradovich	2007/0299824	A1	12/2007	Pan et al.	707/3
2005/0234727	A1	10/2005	Chiu	2008/0014908	A1	1/2008	Vasant	
2005/0246174	A1	11/2005	DeGolia	2008/0034032	A1	2/2008	Healey et al.	709/203
2005/0283364	A1 *	12/2005	Longe	2008/0046311	A1	2/2008	Shahine	
				2008/0059188	A1	3/2008	Konopka	
2005/0283532	A1 *	12/2005	Kim	2008/0065386	A1	3/2008	Cross et al.	704/270
				2008/0065389	A1	3/2008	Cross	
				2008/0065390	A1 *	3/2008	Ativanichayaphong	G06F 9/4446
								704/270
2005/0283752	A1	12/2005	Fruchter et al.	2008/0086455	A1	4/2008	Meisels	
2006/0041431	A1	2/2006	Maes	2008/0091406	A1	4/2008	Baldwin et al.	704/4
2006/0046740	A1	3/2006	Johnson	2008/0103761	A1	5/2008	Printz et al.	704/9
2006/0047509	A1	3/2006	Ding et al.	2008/0103781	A1	5/2008	Wasson	
2006/0072738	A1	4/2006	Louis et al.	2008/0104071	A1	5/2008	Pragada	
2006/0074670	A1 *	4/2006	Weng	2008/0109285	A1	5/2008	Reuther et al.	705/7
				2008/0115163	A1	5/2008	Gilboa et al.	725/34
2006/0074671	A1 *	4/2006	Farmaner	2008/0126091	A1	5/2008	Clark	
				2008/0126284	A1	5/2008	Forbes	
2006/0080098	A1 *	4/2006	Campbell	2008/0133215	A1	6/2008	Sarukkai	704/2
				2008/0140385	A1	6/2008	Mahajan et al.	704/9
				2008/0147396	A1	6/2008	Wang	704/246
2006/0100851	A1	5/2006	Schonebeck	2008/0147410	A1	6/2008	Odinak	704/270.1
2006/0106769	A1 *	5/2006	Gibbs	2008/0147637	A1	6/2008	Li	
2006/0129409	A1 *	6/2006	Mizutani	2008/0154604	A1	6/2008	Sathish et al.	704/257
				2008/0162471	A1	7/2008	Bernard	707/5
				2008/0177530	A1	7/2008	Cross et al.	704/4
				2008/0184164	A1 *	7/2008	Di Fabbriozio	G06F 3/167
								715/809
2006/0242017	A1	10/2006	Libes	2008/0189110	A1	8/2008	Freeman et al.	704/251
2006/0253247	A1	11/2006	de Silva	2008/0189187	A1	8/2008	Hao	
2006/0253281	A1	11/2006	Letzt	2008/0228496	A1	9/2008	Yu	
2006/0285662	A1	12/2006	Yin et al.	2008/0235023	A1	9/2008	Kennewick et al.	704/257
2007/0011159	A1 *	1/2007	Hillis	2008/0235027	A1	9/2008	Cross	704/270.1
2007/0033005	A1	2/2007	Cristo et al.	2008/0269958	A1	10/2008	Filev	
2007/0033020	A1	2/2007	(Kelleher) Francois et al.	2008/0270135	A1	10/2008	Goel	
				2008/0270224	A1	10/2008	Portman	
2007/0033526	A1	2/2007	Thompson	2008/0294437	A1	11/2008	Nakano	
2007/0038436	A1	2/2007	Cristo et al.	2008/0294994	A1	11/2008	Kruger	
2007/0038445	A1	2/2007	Helbing et al.	2008/0306743	A1 *	12/2008	Di Fabbriozio	G10L 15/22
2007/0043569	A1	2/2007	Potter, III et al.					704/275
2007/0043574	A1	2/2007	Coffman et al.	2008/0319751	A1	12/2008	Kennewick et al.	704/257
2007/0043868	A1	2/2007	Kumar et al.	2009/0006077	A1	1/2009	Keaveney	
2007/0050191	A1	3/2007	Weider et al.	2009/0006194	A1	1/2009	Prashantidharan	
2007/0050279	A1	3/2007	Huang	2009/0018829	A1 *	1/2009	Kuperstein	G10L 15/26
2007/0055525	A1	3/2007	Kennewick et al.					704/235
2007/0060114	A1	3/2007	Ramer	2009/0024476	A1	1/2009	Baar	
2007/0061067	A1	3/2007	Zeinstra et al.	2009/0030686	A1	1/2009	Weng	
2007/0061735	A1	3/2007	Hoffberg et al.	2009/0052635	A1	2/2009	Jones et al.	379/88.02
2007/0067310	A1	3/2007	Gupta	2009/0055176	A1	2/2009	Hu	
2007/0073544	A1	3/2007	Millett et al.	2009/0067599	A1	3/2009	Agarwal et al.	379/114.21
2007/0078708	A1	4/2007	Yu et al.	2009/0076827	A1	3/2009	Bulitta et al.	704/275
2007/0078709	A1	4/2007	Rajaram	2009/0106029	A1	4/2009	DeLine et al.	704/275
2007/0078814	A1	4/2007	Flowers	2009/0117885	A1	5/2009	Roth	455/414.3
2007/0094003	A1	4/2007	Huang et al.	2009/0144131	A1	6/2009	Chiu	
2007/0100797	A1 *	5/2007	Thun	2009/0144271	A1	6/2009	Richardson et al.	707/5
2007/0106499	A1	5/2007	Dahlgren	2009/0150156	A1	6/2009	Kennewick et al.	704/257
2007/0112555	A1	5/2007	Lavi	2009/0157382	A1	6/2009	Bar	
2007/0112630	A1	5/2007	Lau	2009/0164216	A1	6/2009	Chengalvarayan	
2007/0118357	A1	5/2007	Kasravi et al.	2009/0171664	A1	7/2009	Kennewick et al.	704/257
2007/0124057	A1	5/2007	Prieto	2009/0171912	A1	7/2009	Nash	
2007/0135101	A1	6/2007	Ramati et al.	2009/0197582	A1	8/2009	Lewis	
2007/0146833	A1	6/2007	Satomi et al.	2009/0216540	A1	8/2009	Tessel et al.	704/275
2007/0162296	A1	7/2007	Altberg et al.	2009/0248565	A1	10/2009	Chuang	
2007/0174258	A1	7/2007	Jones					

US 10,297,249 B2

Page 8

(56)

References Cited

U.S. PATENT DOCUMENTS

2009/0248605	A1	10/2009	Mitchell		2014/0012577	A1	1/2014	Freeman et al.	704/249
2009/0259561	A1	10/2009	Boys		2014/0025371	A1	1/2014	Min	
2009/0259646	A1	10/2009	Fujita et al.	707/5	2014/0108013	A1	4/2014	Di Cristo et al.	704/254
2009/0265163	A1	10/2009	Li		2014/0136187	A1	5/2014	Wolverton	
2009/0271194	A1	10/2009	Davis et al.	704/235	2014/0156278	A1	6/2014	Kennewick et al.	704/254
2009/0273563	A1	11/2009	Pryor	345/157	2014/0195238	A1	7/2014	Terao	
2009/0276700	A1	11/2009	Anderson et al.	715/700	2014/0236575	A1	8/2014	Tur	
2009/0287680	A1	11/2009	Paek		2014/0249821	A1	9/2014	Kennewick et al.	704/257
2009/0299745	A1	12/2009	Kennewick et al.	704/257	2014/0249822	A1	9/2014	Baldwin et al.	704/257
2009/0299857	A1	12/2009	Brubaker		2014/0278413	A1	9/2014	Pitschel	
2009/0304161	A1	12/2009	Pettyjohn		2014/0278416	A1	9/2014	Schuster	
2009/0307031	A1	12/2009	Winkler et al.	705/7	2014/0288934	A1	9/2014	Kennewick et al.	704/254
2009/0313026	A1	12/2009	Coffman et al.	704/275	2014/0330552	A1	11/2014	Bangalore	
2009/0319517	A1	12/2009	Guha		2014/0337007	A1	11/2014	Waibel	
2010/0023320	A1	1/2010	Di Cristo et al.	704/9	2014/0365222	A1	12/2014	Weider et al.	704/257
2010/0023331	A1	1/2010	Duta		2015/0019211	A1	1/2015	Simard	
2010/0029261	A1	2/2010	Mikkelsen et al.	455/419	2015/0019217	A1	1/2015	Di Cristo et al.	704/235
2010/0036967	A1	2/2010	Caine et al.	709/236	2015/0019227	A1	1/2015	Anandarajah	
2010/0049501	A1	2/2010	Kennewick et al.	704/9	2015/0066479	A1	3/2015	Pasupalak	
2010/0049514	A1	2/2010	Kennewick et al.	704/233	2015/0066627	A1	3/2015	Freeman et al.	705/14.41
2010/0057443	A1	3/2010	Di Cristo et al.	704/9	2015/0073910	A1	3/2015	Kennewick et al.	705/14.58
2010/0063880	A1	3/2010	Atsmon et al.	705/14.53	2015/0095159	A1	4/2015	Kennewick et al.	705/14.58
2010/0064025	A1	3/2010	Nelimarkka		2015/0142447	A1	5/2015	Kennewick	
2010/0094707	A1	4/2010	Freer	705/14.54	2015/0170641	A1	6/2015	Kennewick	
2010/0138300	A1	6/2010	Wallis		2015/0193379	A1	7/2015	Mehta	
2010/0145700	A1	6/2010	Kennewick et al.	704/257	2015/0199339	A1	7/2015	Mirkin	
2010/0185512	A1	7/2010	Borger et al.	705/14.49	2015/0293917	A1	10/2015	Bufe	
2010/0191856	A1	7/2010	Gupta		2015/0348544	A1	12/2015	Baldwin	
2010/0204986	A1	8/2010	Kennewick et al.	704/226	2015/0348551	A1	12/2015	Gruber	
2010/0204994	A1	8/2010	Kennewick et al.	704/257	2015/0364133	A1	12/2015	Freeman	
2010/0217604	A1	8/2010	Baldwin et al.	704/275	2016/0049152	A1	2/2016	Kennewick	
2010/0268536	A1	10/2010	Suendermann		2016/0078482	A1	3/2016	Kennewick	
2010/0286985	A1	11/2010	Kennewick et al.	704/257	2016/0078491	A1	3/2016	Kennewick	
2010/0299142	A1	11/2010	Freeman et al.	704/9	2016/0078504	A1	3/2016	Kennewick	
2010/0312547	A1	12/2010	Van Os		2016/0078773	A1	3/2016	Carter	
2010/0312566	A1	12/2010	Odinak et al.	704/275	2016/0110347	A1	4/2016	Kennewick	
2010/0318357	A1	12/2010	Istvan		2016/0148610	A1	5/2016	Kennewick	
2010/0331064	A1	12/2010	Michelstein		2016/0148612	A1	5/2016	Guo	
2011/0022393	A1	1/2011	Waller et al.	704/270	2016/0188292	A1	6/2016	Carter	
2011/0106527	A1	5/2011	Chiu		2016/0188573	A1	6/2016	Tang	
2011/0112827	A1	5/2011	Kennewick et al.	704/9	2016/0217785	A1	7/2016	Kennewick	
2011/0112921	A1	5/2011	Kennewick et al.	705/26.1	2016/0335676	A1	11/2016	Freeman	
2011/0119049	A1	5/2011	Ylonen						
2011/0131036	A1	6/2011	Di Cristo et al.	704/9					
2011/0131045	A1	6/2011	Cristo et al.	704/249					
2011/0231182	A1	9/2011	Weider et al.	704/9					
2011/0231188	A1	9/2011	Kennewick et al.	704/236					
2011/0238409	A1*	9/2011	Larcheveque	G06F 17/2785 704/9					
2011/0307167	A1	12/2011	Taschereau						
2012/0022857	A1	1/2012	Baldwin et al.	704/9					
2012/0041753	A1	2/2012	Dymetman						
2012/0046935	A1	2/2012	Nagao						
2012/0101809	A1	4/2012	Kennewick et al.	704/9					
2012/0101810	A1	4/2012	Kennewick et al.	704/9					
2012/0109753	A1	5/2012	Kennewick et al.	705/14.58					
2012/0150620	A1	6/2012	Mandyam						
2012/0150636	A1	6/2012	Freeman et al.	705/14.49					
2012/0239498	A1	9/2012	Ramer						
2012/0240060	A1	9/2012	Pennington						
2012/0265528	A1	10/2012	Gruber						
2012/0278073	A1	11/2012	Weider et al.	704/235					
2013/0006734	A1	1/2013	Ocko						
2013/0054228	A1	2/2013	Baldwin et al.	704/9					
2013/0060625	A1	3/2013	Davis						
2013/0080177	A1	3/2013	Chen	704/275					
2013/0211710	A1	8/2013	Kennewick et al.	701/419					
2013/0253929	A1	9/2013	Weider et al.	704/235					
2013/0254314	A1	9/2013	Chow						
2013/0297293	A1	11/2013	Di Cristo et al.	704/9					
2013/0304473	A1	11/2013	Baldwin et al.	704/257					
2013/0311324	A1	11/2013	Stoll						
2013/0332454	A1	12/2013	Stuhec						
2013/0339022	A1	12/2013	Baldwin et al.	704/257					
2014/0006951	A1	1/2014	Hunter						

FOREIGN PATENT DOCUMENTS

EP	1 320 043	A2	6/2003
EP	1 646 037		4/2006
JP	H08263258		10/1996
JP	H11249773		9/1999
JP	2001071289		3/2001
JP	2006-146881		6/2006
JP	2008-027454		2/2008
JP	2008058465		3/2008
JP	2008-139928		6/2008
JP	2011504304		2/2011
JP	2012518847		8/2012
WO	WO 99/46763		9/1999
WO	WO 00/21232		4/2000
WO	WO 00/46792		8/2000
WO	0129742		4/2001
WO	0171609	A2	9/2001
WO	WO 01/78065		10/2001
WO	WO 2004/072954		8/2004
WO	2005010702	A2	2/2005
WO	WO 2007/019318		2/2007
WO	WO 2007/021587		2/2007
WO	WO 2007/027546		3/2007
WO	WO 2007/027989		3/2007
WO	WO 2008/098039		8/2008
WO	WO 2008/118195		10/2008
WO	WO 2009/075912		6/2009
WO	2009111721		9/2009
WO	WO 2009/145796		12/2009
WO	WO 2010/096752		8/2010
WO	2016044290		3/2016
WO	2016044316		3/2016
WO	2016044319		3/2016

US 10,297,249 B2

Page 9

(56)

References Cited

FOREIGN PATENT DOCUMENTS

WO	2016044321	3/2016
WO	2016061309	4/2016

OTHER PUBLICATIONS

Wu, Su-Lin, et al., "Incorporating Information from Syllable-Length Time Scales into Automatic Speech Recognition", Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, vol. 2, IEEE, 1998, 4 pages.

Wu, Su-Lin, et al., "Integrating Syllable Boundary Information into Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, 1997, vol. 2, IEEE, 1997, 4 pages.

Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., Jul. 28, 2002, 1 page.

Lin, Bor-shen, et al., "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History", ASRU'99, 1999, 4 pages.

Kuhn, Thomas, et al., "Hybrid In-Car Speech Recognition for Mobile Multimedia Applications", Vehicular Technology Conference, IEEE, Jul. 1999, pp. 2009-2013.

Belvin, Robert, et al., "Development of the HRL Route Navigation Dialogue System", Proceedings of the First International Conference on Human Language Technology Research, San Diego, 2001, pp. 1-5.

Lind, R., et al., "The Network Vehicle—A Glimpse into the Future of Mobile Multi-Media" *IEEE Aerosp. Electron. Systems Magazine*, vol. 14, No. 9, Sep. 1999, pp. 27-32.

Zhao, Yilin, "Telematics: Safe and Fun Driving", *IEEE Intelligent Systems*, vol. 17, Issue 1, 2002, pp. 10-14.

Chai et al., "MIND: A Semantics-Based Multimodal Interpretation Framework for Conversational System", *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, Jun. 2002, pp. 37-46.

Cheyen et al., "Multimodal Maps: An Agent-Based Approach", *International Conference on Cooperative Multimodal Communication* (CMC/95), May 24-26, 1995, pp. 111-121.

Elio et al., "On Abstract Task Models and Conversation Policies" in Workshop on Specifying and Implementing Conversation Policies, *Autonomous Agents '99*, Seattle, 1999, 10 pages.

Turunen, "Adaptive Interaction Methods in Speech User Interfaces", Conference on Human Factors in Computing Systems, Seattle, Washington, 2001, pp. 91-92.

Mao, Mark Z., "Automatic Training Set Segmentation for Multi-Pass Speech Recognition", Department of Electrical Engineering, Stanford University, CA, copyright 2005, IEEE, pp. I-685 to I-688.

Vanhoecke, Vincent, "Confidence Scoring and Rejection Using Multi-Pass Speech Recognition", Nuance Communications, Menlo Park, CA, 2005, 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, 1998, 4 pages.

El Meliani et al., "A Syllabic-Filler-Based Continuous Speech Recognizer for Unlimited Vocabulary", Canadian Conference on Electrical and Computer Engineering, vol. 2, Sep. 5-8, 1995, pp. 1007-1010.

Arrington, Michael, "Google Redefines GPS Navigation Landscape: Google Maps Navigation for Android 2.0", TechCrunch, printed from the Internet <<http://www.techcrunch.com/2009/10/28/google-redefines-car-gps-navigation-google-maps-navigation-android/>>, Oct. 28, 2009, 4 pages.

Bazzi, Issam et al., "Heterogeneous Lexical Units for Automatic Speech Recognition: Preliminary Investigations", *Processing of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Jun. 5-9, 2000, XP010507574, pp. 1257-1260.

O'Shaughnessy, Douglas, "Interacting with Computers by Voice: Automatic Speech Recognition and Synthesis", *Proceedings of the IEEE*, vol. 91, No. 9, Sep. 1, 2003, XP011100665, pp. 1272-1305.

Statement in Accordance with the Notice from the European Patent Office dated Oct. 1, 2007 Concerning Business Methods (OJ EPO Nov. 2007, 592-593), XP002456252.

Davis, Z., et al., A Personal Handheld Multi-Modal Shopping Assistant, IEEE, 2006, 9 pages.

* cited by examiner

U.S. Patent

May 21, 2019

Sheet 1 of 3

US 10,297,249 B2

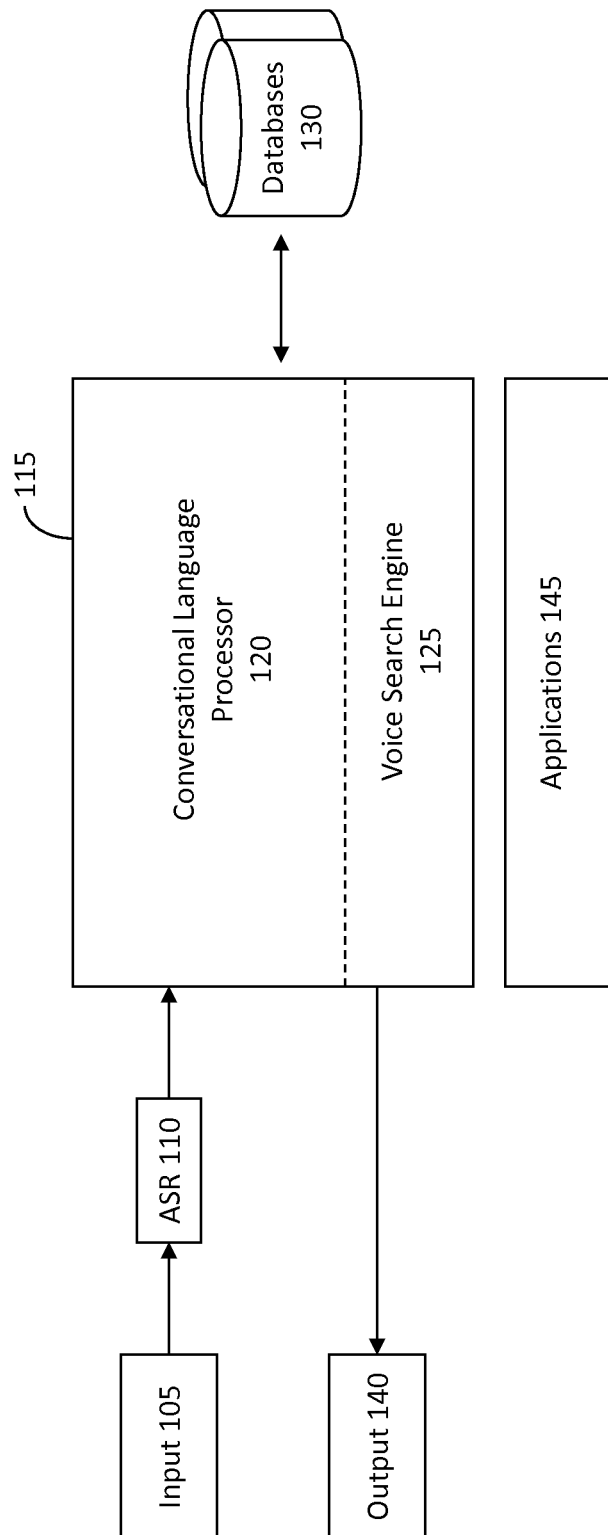


Figure 1

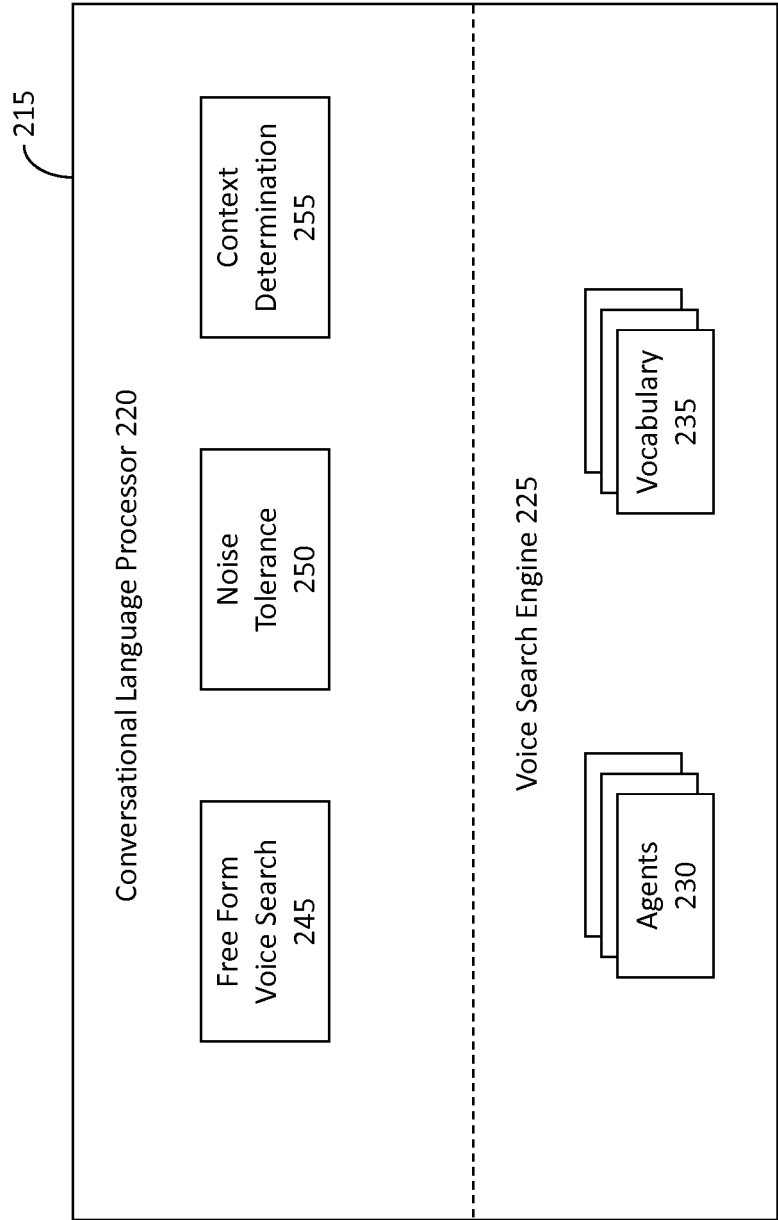


Figure 2

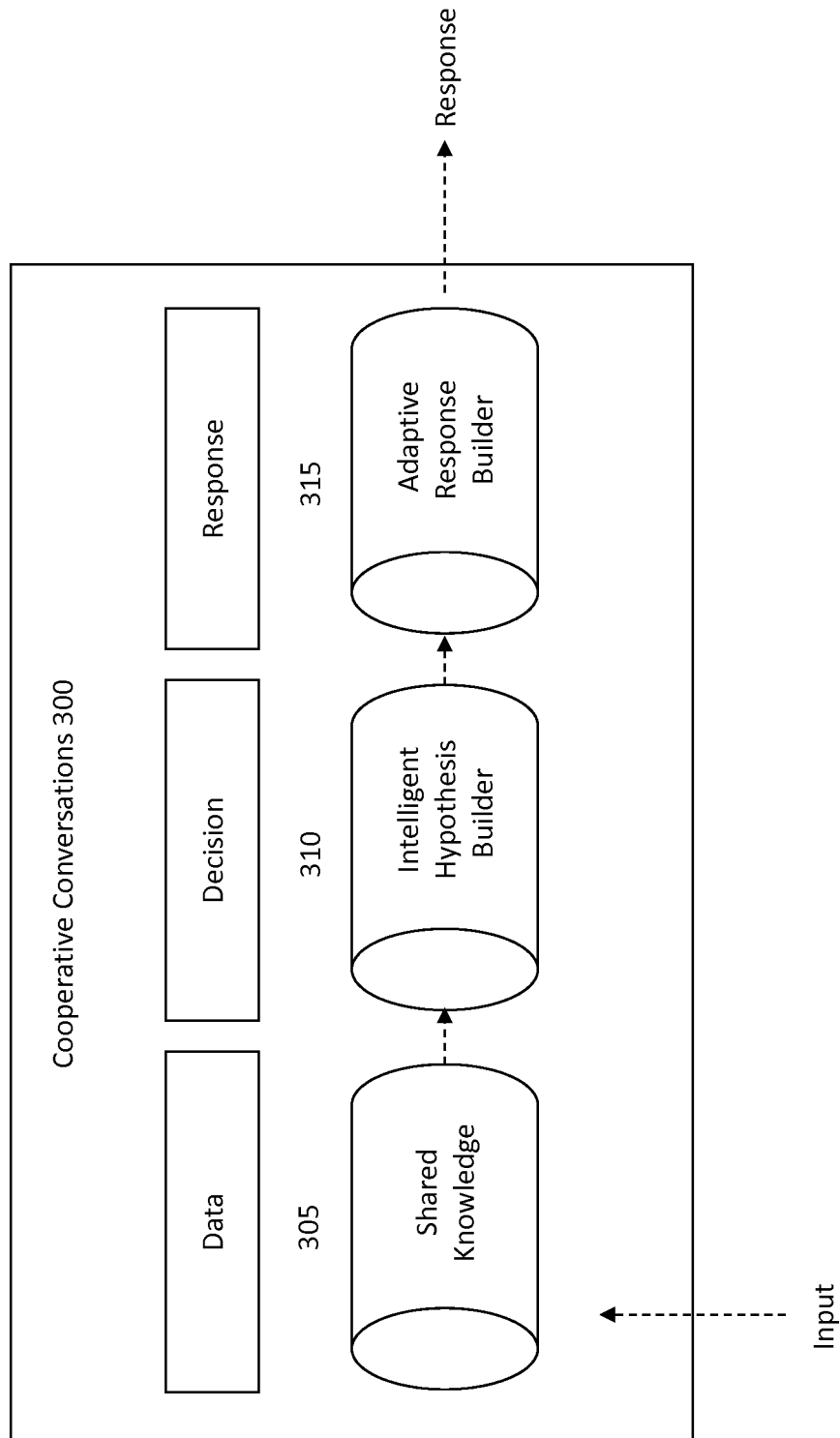


Figure 3

US 10,297,249 B2

1

SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. patent application Ser. No. 13/987,645, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed Aug. 19, 2013, which is a divisional of U.S. patent application Ser. No. 13/251,712, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed Oct. 3, 2011 (which issued as U.S. Pat. No. 8,515,765 on Aug. 20, 2013), which is a continuation of U.S. patent application Ser. No. 11/580,926, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed Oct. 16, 2006 (which issued as U.S. Pat. No. 8,073,681 on Dec. 6, 2011), each of which are hereby incorporated by reference in its entirety.

FIELD OF THE INVENTION

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

BACKGROUND OF THE INVENTION

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.

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

2

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.

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

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.

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 a speech recognition engine (which may alternatively be referred to as an Automatic Speech Recognizer or ASR) 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.

According to another aspect of the invention, an exemplary conversational 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

US 10,297,249 B2

3

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

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.

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 agent “wins” the competition, the winning domain agent may be

4

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

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.

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 conversational model 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. 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 knowl-

US 10,297,249 B2

5

edge) to inform a decision (intelligent hypothesis building), and then may refine the decision and generate a response (adaptive response building).

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 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.

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 dynami-

6

cally as information is reallocated from one participant to another, or as conversational goals change based on the reallocation of information).

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.

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 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).

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."

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.

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 inter-

US 10,297,249 B2

7

faces, and users may respond by using whichever input method, or combination of input methods, is most convenient.

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.

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

FIG. 1 is an exemplary block diagram of a system architecture according to one aspect of the invention.

FIG. 2 is an exemplary block diagram of a conversational speech engine according to one aspect of the invention.

FIG. 3 is an exemplary block diagram of a cooperative conversational model according to one aspect of the invention.

DETAILED DESCRIPTION

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

The utterance component of input **105** may be processed by a speech recognition engine **110** (which may alternatively be referred to herein as Automatic Speech Recognizer **110**, or as shown in FIG. 1, ASR **110**) to generate one or more preliminary interpretations of the utterance. The speech recognition engine **110** may process the utterance using any suitable technique known in the art. For example, in one implementation, the speech recognition engine **110** may interpret the utterance using techniques of phonetic dictation to recognize a phoneme stream, as described in U.S. patent application Ser. No. 11/513,269, entitled “Dynamic Speech Sharpening,” filed Aug. 31, 2006, which issued as U.S. Pat. No. 7,634,409 on Dec. 15, 2009, and which is hereby incorporated by reference in its entirety. The one or more preliminary interpretations generated by the speech recognition engine **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

8

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, the speech recognition engine **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).

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 domain agents **230** and/or one or more vocabularies **235** to aid in interpreting utterances and generating responses, as described in “Enhancing the VUE™ (Voice-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.

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.

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).

US 10,297,249 B2

9

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 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).

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.

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. 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

10

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.

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.

The imperfect speech feature may be able to infer requests 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, 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.

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

US 10,297,249 B2

11

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 those conditions.

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 U.S. patent application Ser. No. 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," filed Aug. 5, 2005, which issued as U.S. Pat. No. 7,640,160 on Dec. 29, 2009 and U.S. patent application Ser. No. 11/212,693, entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," filed Aug. 29, 2005, which issued as U.S. Pat. No. 7,949,529 on May 24, 2011, 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).

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 grammars associated with the speech recognition engine 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.

12

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

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, Ore. based on the Weather and an environmental proximity between Portland, Ore. and Seattle, Wash. 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, N.H.," in the second utterance, context determination process 255 may instead retrieve a weather report for Portland, Me. 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 disambiguate between possibilities, responses may prompt the user with a request for further information (e.g., "Did you mean Portland, Me., or Portland, Ore.?).

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

US 10,297,249 B2

13

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.

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 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.

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

14

building) **310**, and then may refine the decision and generate a response (adaptive response building) **315**.

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.

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

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 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.

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,

US 10,297,249 B2

15

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.

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).

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 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:

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

16

-continued

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

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

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.

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

US 10,297,249 B2

17

needed to complete a request or task, data in the Session Input Accumulator may be relied upon to tune a degree of certainty.

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, 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.

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 a confidence level associated with a preliminary interpretation generated at the speech recognition engine 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 confidence level associated with the preliminary interpretation generated at the speech recognition engine 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 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

18

with a movie context domain agent so that responses may occasionally include French movies).

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.

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.

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 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?

US 10,297,249 B2

19

-continued

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

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

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.

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.

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

20

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:

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

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.

What is claimed is:

1. A computer-implemented method of facilitating natural language system responses using short-term knowledge generated based on one or more prior multi-modal device interactions, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:

receiving, by the computer system during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;

receiving, by the computer system, a second voice input comprising the first natural language utterance via a second input device;

comparing, by the computer system, the first voice input with the second voice input;

filtering, by the computer system, sound from the first voice input and the second voice input based on the comparison;

obtaining, by the computer system during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input;

generating, by the computer system, the short-term knowledge based on at least the first voice input and the first non-voice input;

determining, by the computer system, based on the short-term knowledge, a first context for the first natural language utterance;

determining, by the computer system, based on the first context, an interpretation of the first natural language utterance; and

generating, by the computer system, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.

2. The method of claim 1, the method further comprising: receiving, at the computer system, a second natural language utterance via the first input device;

US 10,297,249 B2

21

determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to the first context;

determining, by the computer system, responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge;

determining, by the computer system, based on the second context, an interpretation of the second natural language utterance; and

generating, by the computer system, based on the interpretation of the second natural language utterance, a response to the second natural language utterance.

3. The method of claim 1, the method further comprising: tracking, by the computer system, contexts identified for multiple consecutive natural language user utterances to generate a context stack, the context stack including the contexts identified for the multiple consecutive natural language user utterances;

receiving, at the computer system, a second natural language utterance via the first input device; and

determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.

4. The method of claim 3, wherein determining whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack includes:

determining whether the second natural language utterance corresponds to a most recent context included in the context stack; and

responsive to the second natural language utterance not corresponding to the most recent context, determining whether the second natural language utterance corresponds to a second most recent context included in the context stack.

5. The method of claim 3, further comprising:

determining, by the computer system, responsive to the second natural language utterance not corresponding to one or more individual ones of the contexts included in the context stack, a second context for the second natural language utterance based on the short-term knowledge; and

determining, by the computer system, based on the second context, an interpretation of the second natural language utterance.

6. The method of claim 1, wherein determining the interpretation of the first natural language utterance comprises determining, based on the first context, an interpretation of one or more recognized words of the first natural language utterance.

7. The method of claim 1, the method further comprising: expiring, by the computer system, one or more items of short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation.

8. The method of claim 7, further comprising: accumulating, by the computer system, long-term knowledge based on the one or more items of short-term knowledge that are expired.

9. The method of claim 1, the method further comprising: accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received during conversations occurring prior to the

22

first conversation, wherein the first context is determined based on the short-term knowledge and the long-term knowledge.

10. The method of claim 1, further comprising: determining, by the computer system, based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the first response is generated based on the determined manner and the interpretation of the first natural language utterance.

11. The method of claim 10, wherein the manner in which the first natural language utterance is spoken includes at least one of a tone, a pace, an inflection, or a timing.

12. The method of claim 1, wherein the second input device is included with the first input device in a single device.

13. The method of claim 1, wherein generating the first response comprises generating, by the computer system, a non-voice response in a graphical user interface and/or a voice response via a speaker.

14. The method of claim 1, wherein the first natural language utterance relates to a request and the first response is intended to be responsive to the request, the method further comprising:

receiving, by the computer system, a second natural language utterance in response to the first response, wherein the second natural language utterance comprises a clarification related to the interpretation of the first natural language utterance; and

responsive to receipt of the second natural language utterance, generating, by the computer system, a response to the second natural language utterance based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance.

15. The method of claim 1, wherein the first non-voice input is received via a non-voice input device comprising a touch screen device, the method further comprising:

receiving, by the computer system, information indicating a location on the touch screen device that received the first non-voice input.

16. A system for facilitating natural language system responses via short-term knowledge generated based on one or more prior multi-modal device interactions, the system comprising:

one or more physical processors programmed with one or more computer program instructions which, when executed, cause the one or more physical processors to:

receive, during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;

receive a second voice input comprising the first natural language utterance via a second input device;

compare the first voice input with the second voice input;

filter sound from the first voice input and the second voice input based on the comparison;

obtain, during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input;

generate the short-term knowledge based on at least the first voice input and the first non-voice input;

determine, based on the short-term knowledge, a first context for the first natural language utterance;

determine, based on the first context, an interpretation of the first natural language utterance; and generate,

US 10,297,249 B2

23

based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.

17. The system of claim 16, wherein the one or more physical processors are further caused to:

receive a second natural language utterance via the first input device;

determine, based on the short-term knowledge, whether the second natural language utterance corresponds to the first context;

determine, responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge;

determine, based on the second context, an interpretation of the second natural language utterance; and

generate, based on the interpretation of the second natural language utterance, a response to the second natural language utterance.

18. The system of claim 16, wherein the one or more physical processors are further caused to:

track contexts identified for multiple consecutive natural language user utterances to generate a context stack, the context stack including the contexts identified for the multiple consecutive natural language user utterances; receive a second natural language utterance via the first input device; and

determine, based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.

19. The system of claim 18, wherein determining whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack includes:

determining whether the second natural language utterance corresponds to a most recent context included in the context stack; and

responsive to the second natural language utterance not corresponding to the most recent context, determining whether the second natural language utterance corresponds to a second most recent context included in the context stack.

20. The system of claim 18, wherein the one or more physical processors are caused to:

determine, responsive to the second natural language utterance not corresponding to one or more individual ones of the contexts included in the context stack, a second context for the second natural language utterance based on the short-term knowledge; and

determine, based on the second context, an interpretation of the second natural language utterance.

21. The system of claim 16, wherein determining the interpretation of the first natural language utterance com-

24

prises determining, based on the first context, an interpretation of one or more recognized words of the first natural language utterance.

22. The system of claim 16, wherein the one or more physical processors are caused to:

expire one or more items of accumulated short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation; and

accumulate long-term knowledge based on the one or more items of accumulated short-term knowledge that are expired, wherein the first context is determined based on short-term knowledge and the long-term knowledge.

23. The system of claim 16, wherein the one or more physical processors are caused to:

determine, based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the first response is generated based on the determined manner and the interpretation of the first natural language utterance.

24. The system of claim 23, wherein the manner in which the first natural language utterance is spoken includes at least one of a tone, a pace, an inflection, or a timing.

25. The system of claim 16, wherein the second input device is included with the first input device in a single device.

26. The system of claim 16, wherein to generate the first response, the one or more physical processors are caused to: generate a non-voice response in a graphical user interface and/or a voice response via a speaker.

27. The system of claim 16, wherein the first natural language utterance relates to a request and the first response is intended to be responsive to the request, wherein the one or more physical processors are further caused to:

receive a second natural language utterance in response to the first response, wherein the second natural language utterance comprises a clarification related to the interpretation of the first natural language utterance; and responsive to receipt of the second natural language utterance, generate a response to the second natural language utterance based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance.

28. The system of claim 16, wherein the first non-voice input is received via a non-voice input device comprising a touch screen device, and wherein the one or more physical processors are further caused to:

receive information indicating a location on the touch screen device that received the first non-voice input.

* * * * *

EXHIBIT B



US010755699B2

(12) **United States Patent**
Baldwin et al.

(10) **Patent No.:** **US 10,755,699 B2**

(45) **Date of Patent:** **Aug. 25, 2020**

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

(71) Applicant: **VB Assets, LLC**, Bellevue, WA (US)

(72) Inventors: **Larry Baldwin**, Maple Valley, WA
(US); **Tom Freeman**, Mercer Island,
WA (US); **Michael Tjalve**, Bellevue,
WA (US); **Blane Ebersold**, Seattle, WA
(US); **Chris Weider**, Seattle, WA (US)

(73) Assignee: **VB Assets, LLC**, Bellevue, WA (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

(21) Appl. No.: **16/417,178**

(22) Filed: **May 20, 2019**

(65) **Prior Publication Data**

US 2019/0272823 A1 Sep. 5, 2019

Related U.S. Application Data

(60) Division of application No. 14/691,445, filed on Apr.
20, 2015, now Pat. No. 10,297,249, which is a
(Continued)

(51) **Int. Cl.**
G10L 15/18 (2013.01)
G10L 25/51 (2013.01)
(Continued)

(52) **U.S. Cl.**
CPC **G10L 15/18** (2013.01); **G06F 3/167**
(2013.01); **G06F 40/30** (2020.01);
(Continued)

(58) **Field of Classification Search**
CPC G10L 15/1815; G10L 15/1822; G10L
15/183; G10L 15/22; G10L 2015/225;
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,430,669 A 2/1984 Cheung
4,821,027 A 4/1989 Mallory
(Continued)

FOREIGN PATENT DOCUMENTS

CN 1433554 A 7/2003
CN 1860496 A 11/2006
(Continued)

OTHER PUBLICATIONS

"Statement in Accordance with the Notice from the European Patent
Office" dated Oct. 1, 2007 Concerning Business Methods (OJ EPO
Nov. 2007, 592-593), XP002456252.

(Continued)

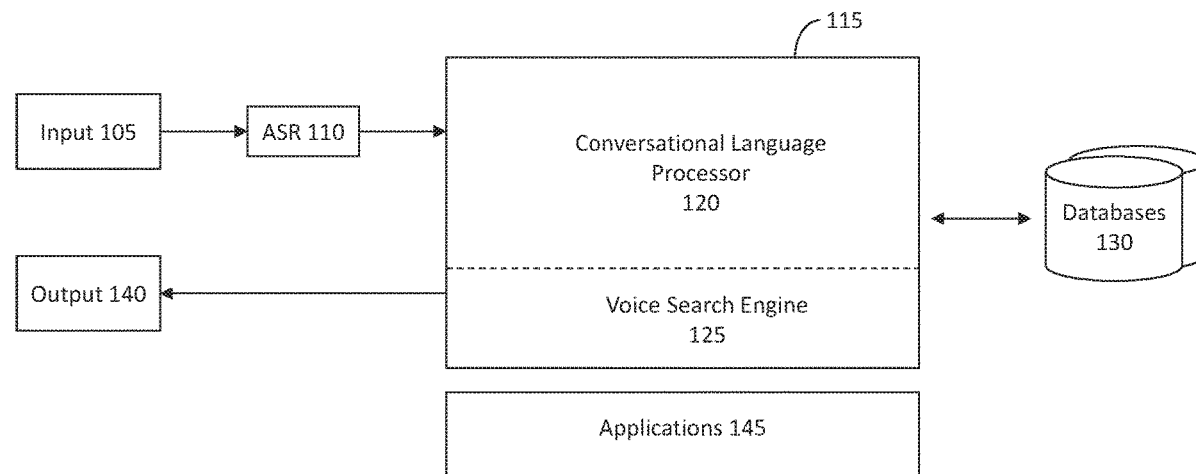
Primary Examiner — Eric Yen

(74) *Attorney, Agent, or Firm* — Sheppard Mullin Richter
& Hampton LLP

(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.

22 Claims, 3 Drawing Sheets



US 10,755,699 B2

Page 2

Related U.S. Application Data					
continuation of application No. 13/987,645, filed on Aug. 19, 2013, now Pat. No. 9,015,049, which is a division of application No. 13/251,712, filed on Oct. 3, 2011, now Pat. No. 8,515,765, which is a continuation of application No. 11/580,926, filed on Oct. 16, 2006, now Pat. No. 8,073,681.			5,708,422 A	1/1998	Blonder
			5,721,938 A	2/1998	Stuckey
			5,722,084 A	2/1998	Chakrin
			5,740,256 A	4/1998	CastelloDaCosta
			5,742,763 A	4/1998	Jones
			5,748,841 A *	5/1998	Morin G06F 3/16 704/257
			5,748,974 A	5/1998	Johnson
			5,752,052 A	5/1998	Richardson
			5,754,784 A	5/1998	Garland
			5,761,631 A	6/1998	Nasukawa
			5,774,841 A	6/1998	Salazar
			5,774,859 A	6/1998	Houser
			5,794,050 A	8/1998	Dahlgren
			5,794,196 A	8/1998	Yegnanarayanan
			5,797,112 A	8/1998	Komatsu
			5,799,276 A	8/1998	Komissarchik
			5,802,510 A	9/1998	Jones
			5,829,000 A	10/1998	Huang
			5,832,221 A	11/1998	Jones
			5,839,107 A	11/1998	Gupta
			5,848,396 A	12/1998	Gerace
			5,855,000 A	12/1998	Waibel
			5,860,059 A	1/1999	Aust
			5,867,817 A	2/1999	Catallo
			5,878,385 A	3/1999	Bralich
			5,878,386 A	3/1999	Coughlin
			5,892,813 A	4/1999	Morin
			5,892,900 A	4/1999	Ginter
			5,895,464 A	4/1999	Bhandari
			5,895,466 A	4/1999	Goldberg
			5,897,613 A	4/1999	Chan
			5,898,760 A	4/1999	Smets
			5,899,991 A	5/1999	Karch
			5,902,347 A	5/1999	Backman
			5,911,120 A	6/1999	Jarett
			5,918,222 A	6/1999	Fukui
			5,926,784 A	7/1999	Richardson
			5,933,822 A	8/1999	Braden-Harder
			5,950,167 A	9/1999	Yaker
			5,953,393 A	9/1999	Culbreth
			5,960,384 A	9/1999	Brash
			5,960,397 A	9/1999	Rahim
			5,960,399 A	9/1999	Barclay
			5,960,447 A	9/1999	Holt
			5,963,894 A	10/1999	Richardson
			5,963,940 A	10/1999	Liddy
			5,982,906 A	11/1999	Ono
			5,983,190 A	11/1999	Trower
			5,987,404 A	11/1999	DellaPietra
			5,991,721 A	11/1999	Asano
			5,995,119 A	11/1999	Cosatto
			5,995,928 A	11/1999	Nguyen
			5,995,943 A	11/1999	Bull
			6,009,382 A	12/1999	Martino
			6,014,559 A	1/2000	Amin
			6,018,708 A	1/2000	Dahan
			6,021,384 A	2/2000	Gorin
			6,028,514 A	2/2000	Lemelson
			6,035,267 A	3/2000	Watanabe
			6,044,347 A	3/2000	Abella
			6,049,602 A	4/2000	Foladare
			6,049,607 A	4/2000	Marash
			6,058,187 A	5/2000	Chen
			6,067,513 A	5/2000	Ishimitsu
			6,073,098 A	6/2000	Buchsbaum
			6,076,057 A	6/2000	Narayanan
			6,076,059 A	6/2000	Glickman
			6,078,886 A	6/2000	Dragosh
			6,081,774 A	6/2000	deHita
			6,081,779 A *	6/2000	Besling G10L 15/065 704/240
			6,085,186 A	7/2000	Christianson
			6,101,241 A	8/2000	Boyce
			6,108,631 A	8/2000	Ruhl
			6,119,087 A	9/2000	Kuhn
			6,119,101 A	9/2000	Peckover
			6,122,613 A	9/2000	Baker
			6,134,235 A	10/2000	Goldman

(51) Int. Cl.

G10L 15/22 (2006.01)

G06F 40/30 (2020.01)

G10L 25/63 (2013.01)

G06F 3/16 (2006.01)

G10L 17/22 (2013.01)

G10L 21/0216 (2013.01)

G10L 15/183 (2013.01)

G10L 15/06 (2013.01)

(52) U.S. Cl.

CPC G10L 15/1815 (2013.01); G10L 15/1822 (2013.01); G10L 15/22 (2013.01); G10L 17/22 (2013.01); G10L 25/51 (2013.01); G10L 25/63 (2013.01); G10L 15/183 (2013.01); G10L 2015/0631 (2013.01); G10L 2015/225 (2013.01); G10L 2015/228 (2013.01); G10L 2021/02166 (2013.01)

(58) Field of Classification Search

CPC G10L 2015/228; G10L 2021/02166; G10L 25/51

See application file for complete search history.

(56)

References Cited

U.S. PATENT DOCUMENTS

4,829,423 A 5/1989 Tennant
4,887,212 A 12/1989 Zamora
4,910,784 A 3/1990 Doddington
5,027,406 A 6/1991 Roberts
5,155,743 A 10/1992 Jacobs
5,164,904 A 11/1992 Sumner
5,208,748 A 5/1993 Flores
5,265,065 A 11/1993 Turtle
5,274,560 A 12/1993 LaRue
5,331,554 A 7/1994 Graham
5,357,596 A 10/1994 Takebayashi
5,369,575 A 11/1994 Lamberti
5,377,350 A 12/1994 Skinner
5,386,556 A 1/1995 Hedin
5,424,947 A 6/1995 Nagao
5,471,318 A 11/1995 Ahuja
5,475,733 A 12/1995 Eisdorfer
5,479,563 A 12/1995 Yamaguchi
5,488,652 A 1/1996 Bielby
5,499,289 A 3/1996 Bruno
5,500,920 A 3/1996 Kupiec
5,517,560 A 5/1996 Greenspan
5,533,108 A 7/1996 Harris
5,537,436 A 7/1996 Bottoms
5,539,744 A 7/1996 Chu
5,557,667 A 9/1996 Bruno
5,559,864 A 9/1996 Kennedy, Jr.
5,563,937 A 10/1996 Bruno
5,577,165 A 11/1996 Takebayashi
5,590,039 A 12/1996 Ikeda
5,608,635 A 3/1997 Tamai
5,615,296 A 3/1997 Stanford
5,617,407 A 4/1997 Bareis
5,633,922 A 5/1997 August
5,634,086 A 5/1997 Ritschev
5,652,570 A 7/1997 Lepkofker
5,675,629 A 10/1997 Raffel
5,696,965 A 12/1997 Dedrick

US 10,755,699 B2

Page 3

(56)

References Cited

U.S. PATENT DOCUMENTS

6,144,667	A	11/2000	Doshi	6,453,292	B2	9/2002	Ramaswamy
6,144,938	A	11/2000	Surace	6,456,711	B1	9/2002	Cheung
6,154,526	A	11/2000	Dahlke	6,456,974	B1	9/2002	Baker
6,160,883	A	12/2000	Jackson	6,466,654	B1	10/2002	Cooper
6,167,377	A	12/2000	Gillick	6,466,899	B1	10/2002	Yano
6,173,266	B1	1/2001	Marx	6,470,315	B1	10/2002	Netsch
6,173,279	B1	1/2001	Levin	6,487,494	B2	11/2002	Odinak
6,175,858	B1	1/2001	Bulfer	6,487,495	B1	11/2002	Gale
6,185,535	B1	2/2001	Hedin	6,498,797	B1	12/2002	Anerousis
6,188,982	B1	2/2001	Chiang	6,499,013	B1	12/2002	Weber
6,192,110	B1	2/2001	Abella	6,501,833	B2	12/2002	Phillips
6,192,338	B1	2/2001	Haszto	6,501,834	B1	12/2002	Milewski
6,195,634	B1	2/2001	Dudemaine	6,505,155	B1	1/2003	Vanbuskirk
6,195,651	B1	2/2001	Handel	6,510,417	B1	1/2003	Woods
6,199,043	B1	3/2001	Happ	6,513,006	B2	1/2003	Howard
6,208,964	B1	3/2001	Sabourin	6,522,746	B1	2/2003	Marchok
6,208,972	B1	3/2001	Grant	6,523,061	B1	2/2003	Halverson
6,219,346	B1	4/2001	Maxemchuk	6,532,444	B1	3/2003	Weber
6,219,643	B1	4/2001	Cohen	6,539,348	B1	3/2003	Bond
6,219,645	B1	4/2001	Byers	6,549,629	B2	4/2003	Finn
6,226,612	B1	5/2001	Srenger	6,553,372	B1	4/2003	Brassell
6,233,556	B1	5/2001	Teunen	6,556,970	B1	4/2003	Sasaki
6,233,559	B1	5/2001	Balakrishnan	6,556,973	B1	4/2003	Lewin
6,233,561	B1	5/2001	Junqua	6,560,576	B1	5/2003	Cohen
6,236,968	B1	5/2001	Kanevsky	6,560,590	B1	5/2003	Shwe
6,243,679	B1	6/2001	Mohri	6,567,778	B1	5/2003	ChaoChang
6,246,981	B1	6/2001	Papineni	6,567,797	B1	5/2003	Schuetze
6,246,990	B1	6/2001	Happ	6,567,805	B1	5/2003	Johnson
6,266,636	B1	7/2001	Kosaka	6,570,555	B1	5/2003	Prevost
6,269,336	B1	7/2001	Ladd	6,570,964	B1	5/2003	Murveit
6,272,455	B1	8/2001	Hoshen	6,571,279	B1	5/2003	Herz
6,272,461	B1	8/2001	Meredith	6,574,597	B1	6/2003	Mohri
6,275,231	B1	8/2001	Obradovich	6,574,624	B1	6/2003	Johnson
6,278,377	B1	8/2001	DeLine	6,578,022	B1	6/2003	Foulger
6,278,968	B1	8/2001	Franz	6,581,103	B1	6/2003	Dengler
6,286,002	B1	9/2001	Axaopoulos	6,584,439	B1	6/2003	Geilhufe
6,288,319	B1	9/2001	Catona	6,587,858	B1	7/2003	Strazza
6,292,767	B1	9/2001	Jackson	6,591,185	B1	7/2003	Polidi
6,301,560	B1	10/2001	Masters	6,591,239	B1	7/2003	McCall
6,308,151	B1	10/2001	Smith	6,594,257	B1	7/2003	Doshi
6,311,159	B1	10/2001	VanTichelen	6,594,367	B1	7/2003	Marash
6,314,402	B1	11/2001	Monaco	6,598,018	B1	7/2003	Junqua
6,321,196	B1	11/2001	Franceschi	6,601,026	B2	7/2003	Appelt
6,356,869	B1	3/2002	Chapados	6,601,029	B1	7/2003	Pickering
6,362,748	B1	3/2002	Huang	6,604,075	B1	8/2003	Brown
6,366,882	B1	4/2002	Bijl	6,604,077	B2	8/2003	Dragosh
6,366,886	B1	4/2002	Dragosh	6,606,598	B1	8/2003	Holthouse
6,374,214	B1	4/2002	Friedland	6,611,692	B2	8/2003	Raffel
6,374,226	B1	4/2002	Hunt	6,614,773	B1	9/2003	Maxemchuk
6,377,913	B1	4/2002	Coffman	6,615,172	B1	9/2003	Bennett
6,377,919	B1	4/2002	Burnett	6,622,119	B1	9/2003	Ramaswamy
6,381,535	B1	4/2002	Durocher	6,629,066	B1	9/2003	Jackson
6,385,596	B1	5/2002	Wiser	6,631,346	B1	10/2003	Karaorman
6,385,646	B1	5/2002	Brown	6,631,351	B1	10/2003	Ramachandran
6,389,398	B1	5/2002	Lustgarten	6,633,846	B1	10/2003	Bennett
6,393,403	B1	5/2002	Majaniemi	6,636,790	B1	10/2003	Lightner
6,393,428	B1	5/2002	Miller	6,643,620	B1	11/2003	Contolini
6,397,181	B1	5/2002	Li	6,647,363	B2	11/2003	Claassen
6,404,878	B1	6/2002	Jackson	6,650,747	B1	11/2003	Bala
6,405,170	B1	6/2002	Phillips	6,658,388	B1 *	12/2003	Kleindienst G10L 15/22
6,408,272	B1	6/2002	White				704/270
6,411,810	B1	6/2002	Maxemchuk	6,678,680	B1	1/2004	Woo
6,411,893	B2	6/2002	Ruhl	6,681,206	B1	1/2004	Gorin
6,415,257	B1	7/2002	Junqua	6,691,151	B1	2/2004	Cheyser
6,418,210	B1	7/2002	Sayko	6,701,294	B1	3/2004	Ball
6,420,975	B1	7/2002	DeLine	6,704,396	B2	3/2004	Parolkar
6,429,813	B2	8/2002	Feigen	6,704,576	B1	3/2004	Brachman
6,430,285	B1	8/2002	Bauer	6,704,708	B1	3/2004	Pickering
6,430,531	B1	8/2002	Polish	6,707,421	B1	3/2004	Drury
6,434,523	B1	8/2002	Monaco	6,708,150	B1	3/2004	Hirayama
6,434,524	B1	8/2002	Weber	6,721,001	B1	4/2004	Berstis
6,434,529	B1	8/2002	Walker	6,721,633	B2	4/2004	Funk
6,442,522	B1	8/2002	Carberry	6,721,706	B1	4/2004	Strubbe
6,446,114	B1	9/2002	Bulfer	6,726,636	B2	4/2004	DerGhazarian
6,453,153	B1	9/2002	Bowker	6,732,088	B1	5/2004	Glance
				6,735,592	B1	5/2004	Neumann
				6,739,556	B1	5/2004	Langston
				6,741,931	B1	5/2004	Kohut
				6,742,021	B1	5/2004	Halverson

US 10,755,699 B2

Page 4

(56)

References Cited

U.S. PATENT DOCUMENTS

6,745,161 B1	6/2004	Arnold	7,092,928 B1	8/2006	Elad
6,751,591 B1	6/2004	Gorin	7,107,210 B2	9/2006	Deng
6,751,612 B1	6/2004	Schuetze	7,107,218 B1	9/2006	Preston
6,754,485 B1	6/2004	Obradovich	7,110,951 B1	9/2006	Lemelson
6,754,627 B2	6/2004	Woodward	7,127,395 B1	10/2006	Gorin
6,754,647 B1	6/2004	Tackett	7,127,400 B2	10/2006	Koch
6,757,544 B2	6/2004	Rangarajan	7,130,390 B2	10/2006	Abburri
6,757,718 B1	6/2004	Halverson	7,136,875 B2	11/2006	Anderson
6,785,651 B1	8/2004	Wang	7,137,126 B1	11/2006	Coffman
6,795,808 B1	9/2004	Strubbe	7,143,037 B1	11/2006	Chestnut
6,801,604 B2	10/2004	Maes	7,143,039 B1	11/2006	Stifelman
6,801,893 B1	10/2004	Backfried	7,146,315 B2	12/2006	Balan
6,804,330 B1	10/2004	Jones	7,146,319 B2	12/2006	Hunt
6,810,375 B1	10/2004	Ejerhed	7,149,696 B2	12/2006	Shimizu
6,813,341 B1	11/2004	Mahoney	7,165,028 B2	1/2007	Gong
6,816,830 B1	11/2004	Kempe	7,170,993 B2	1/2007	Anderson
6,823,308 B2	11/2004	Keiller	7,171,291 B2	1/2007	Obradovich
6,829,603 B1	12/2004	Chai	7,174,300 B2	2/2007	Bush
6,832,230 B1	12/2004	Zilliacus	7,177,798 B2	2/2007	Hsu
6,833,848 B1	12/2004	Wolff	7,184,957 B2	2/2007	Brookes
6,850,603 B1	2/2005	Eberle	7,190,770 B2	3/2007	Ando
6,856,990 B2	2/2005	Barile	7,197,069 B2	3/2007	Agazzi
6,865,481 B2	3/2005	Kawazoe	7,197,460 B1	3/2007	Gupta
6,868,380 B2	3/2005	Kroeker	7,203,644 B2	4/2007	Anderson
6,868,385 B1	3/2005	Gerson	7,206,418 B2	4/2007	Yang
6,871,179 B1	3/2005	Kist	7,207,011 B2	4/2007	Mulvey
6,873,837 B1	3/2005	Yoshioka	7,215,941 B2	5/2007	Beckmann
6,877,001 B2	4/2005	Wolf	7,228,276 B2	6/2007	Omote
6,877,134 B1	4/2005	Fuller	7,231,343 B1	6/2007	Treadgold
6,882,970 B1	4/2005	Garner	7,236,923 B1	6/2007	Gupta
6,901,366 B1	5/2005	Kuhn	7,254,482 B2	8/2007	Kawasaki
6,910,003 B1	6/2005	Arnold	7,272,212 B2	9/2007	Eberle
6,912,498 B2	6/2005	Stevens	7,277,854 B2	10/2007	Bennett
6,915,126 B2	7/2005	Mazzara, Jr.	7,283,829 B2	10/2007	Christenson
6,928,614 B1	8/2005	Everhart	7,283,951 B2	10/2007	Marchisio
6,934,756 B2	8/2005	Maes	7,289,606 B2	10/2007	Sibal
6,937,977 B2	8/2005	Gerson	7,299,186 B2	11/2007	Kusunuki
6,937,982 B2	8/2005	Kitaoka	7,301,093 B2	11/2007	Sater
6,941,266 B1	9/2005	Gorin	7,305,381 B1	12/2007	Poppink
6,944,594 B2	9/2005	Busayapongchai	7,321,850 B2	1/2008	Wakita
6,950,821 B2	9/2005	Faybishenko	7,328,155 B2	2/2008	Endo
6,954,755 B2	10/2005	Reisman	7,337,116 B2	2/2008	Charlesworth
6,959,276 B2	10/2005	Droppo	7,340,040 B1	3/2008	Saylor
6,961,700 B2	11/2005	Mitchell	7,366,285 B2	4/2008	Parolkar
6,963,759 B1	11/2005	Gerson	7,366,669 B2	4/2008	Nishitani
6,964,023 B2	11/2005	Maes	7,376,645 B2	5/2008	Bernard
6,968,311 B2	11/2005	Knockeart	7,380,250 B2	5/2008	Schechter
6,973,387 B2	12/2005	Masclat	7,386,443 B1	6/2008	Parthasarathy
6,975,983 B1	12/2005	Fortescue	7,398,209 B2	7/2008	Kennewick
6,975,993 B1	12/2005	Keiller	7,406,421 B2	7/2008	Odinak
6,980,092 B2	12/2005	Turnbull	7,415,100 B2	8/2008	Cooper
6,983,055 B2	1/2006	Luo	7,415,414 B2	8/2008	Azara
6,990,513 B2	1/2006	Belfiore	7,421,393 B1	9/2008	DiFabrizio
6,996,531 B2	2/2006	Korall	7,424,431 B2	9/2008	Greene
7,003,463 B1	2/2006	Maes	7,447,635 B1	11/2008	Konopka
7,016,849 B2	3/2006	Arnold	7,451,088 B1	11/2008	Ehlen
7,020,609 B2	3/2006	Thrift	7,454,368 B2	11/2008	Stillman
7,024,364 B2	4/2006	Guerra	7,454,608 B2	11/2008	Gopalakrishnan
7,027,586 B2	4/2006	Bushey	7,461,059 B2	12/2008	Richardson
7,027,974 B1	4/2006	Busch	7,472,020 B2	12/2008	Brulle-Drews
7,027,975 B1	4/2006	Pazandak	7,472,060 B1	12/2008	Gorin
7,031,908 B1	4/2006	Huang	7,472,075 B2	12/2008	Odinak
7,035,415 B2	4/2006	Belt	7,477,909 B2	1/2009	Roth
7,036,128 B1	4/2006	Julia	7,478,036 B2	1/2009	Shen
7,043,425 B2	5/2006	Pao	7,487,088 B1	2/2009	Gorin
7,054,817 B2	5/2006	Shao	7,487,110 B2	2/2009	Bennett
7,058,890 B2	6/2006	George	7,493,259 B2	2/2009	Jones
7,062,488 B1	6/2006	Reisman	7,493,559 B1	2/2009	Wolff
7,069,220 B2	6/2006	Coffman	7,502,672 B1	3/2009	Kolls
7,072,834 B2	7/2006	Zhou	7,502,730 B2	3/2009	Wang
7,072,888 B1 *	7/2006	Perkins	7,502,738 B2	3/2009	Kennewick
			7,512,906 B1	3/2009	Baier
			7,516,076 B2	4/2009	Walker
			7,529,675 B2	5/2009	Maes
			7,536,297 B2	5/2009	Byrd
			7,536,374 B2	5/2009	Au
			7,542,894 B2	6/2009	Murata
			7,546,382 B2	6/2009	Healey
			7,548,491 B2	6/2009	Macfarlane

G06F 16/9535
707/733

US 10,755,699 B2

Page 5

(56)

References Cited

U.S. PATENT DOCUMENTS

7,552,054 B1	6/2009	Stifelman	8,326,599 B2	12/2012	Tomeh
7,558,730 B2	7/2009	Davis	8,326,627 B2	12/2012	Kennewick
7,574,362 B2	8/2009	Walker	8,326,634 B2	12/2012	DiCristo
7,577,244 B2	8/2009	Taschereau	8,326,637 B2	12/2012	Baldwin
7,606,708 B2	10/2009	Hwang	8,332,224 B2	12/2012	DiCristo
7,606,712 B1	10/2009	Smith	8,340,975 B1	12/2012	Rosenberger
7,620,549 B2	11/2009	DiCristo	8,346,563 B1	1/2013	Hjelm
7,634,409 B2	12/2009	Kennewick	8,370,147 B2	2/2013	Kennewick
7,640,006 B2	12/2009	Portman	8,447,607 B2	5/2013	Weider
7,640,160 B2	12/2009	DiCristo	8,447,651 B1	5/2013	Scholl
7,640,272 B2	12/2009	Mahajan	8,452,598 B2	5/2013	Kennewick
7,672,931 B2	3/2010	Hurst-Hiller	8,503,995 B2	8/2013	Ramer
7,676,365 B2	3/2010	Hwang	8,509,403 B2	8/2013	Chiu
7,676,369 B2	3/2010	Fujimoto	8,515,765 B2	8/2013	Baldwin
7,684,977 B2	3/2010	Morikawa	8,527,274 B2	9/2013	Freeman
7,693,720 B2	4/2010	Kennewick	8,577,671 B1	11/2013	Barve
7,697,673 B2	4/2010	Chiu	8,589,161 B2	11/2013	Kennewick
7,706,616 B2	4/2010	Kristensson	8,612,205 B2	12/2013	Hanneman
7,729,913 B1	6/2010	Lee	8,612,206 B2	12/2013	Chalabi
7,729,916 B2	6/2010	Coffman	8,620,659 B2	12/2013	DiCristo
7,729,918 B2	6/2010	Walker	8,719,005 B1	5/2014	Lee
7,729,920 B2	6/2010	Chaar	8,719,009 B2	5/2014	Baldwin
7,734,287 B2	6/2010	Ying	8,719,026 B2	5/2014	Kennewick
7,748,021 B2	6/2010	Obradovich	8,731,929 B2	5/2014	Kennewick
7,788,084 B2	8/2010	Brun	8,738,380 B2	5/2014	Baldwin
7,792,257 B1	9/2010	Vanier	8,849,652 B2	9/2014	Weider
7,801,731 B2	9/2010	Odinak	8,849,670 B2	9/2014	DiCristo
7,809,570 B2	10/2010	Kennewick	8,849,696 B2	9/2014	Pansari
7,818,176 B2	10/2010	Freeman	8,849,791 B1	9/2014	Hertschuh
7,831,426 B2	11/2010	Bennett	8,886,536 B2	11/2014	Freeman
7,831,433 B1	11/2010	Belvin	8,972,243 B1	3/2015	Strom
7,856,358 B2	12/2010	Ho	8,972,354 B1	3/2015	Telang
7,873,519 B2	1/2011	Bennett	8,983,839 B2	3/2015	Kennewick
7,873,523 B2	1/2011	Potter	9,009,046 B1	4/2015	Stewart
7,873,654 B2	1/2011	Bernard	9,015,049 B2	4/2015	Baldwin
7,881,936 B2	2/2011	Longe	9,037,455 B1	5/2015	Faaborg
7,890,324 B2	2/2011	Bangalore	9,070,366 B1	6/2015	Mathias
7,894,849 B2	2/2011	Kass	9,070,367 B1	6/2015	Hoffmeister
7,902,969 B2	3/2011	Obradovich	9,105,266 B2	8/2015	Baldwin
7,917,367 B2	3/2011	DiCristo	9,171,541 B2	10/2015	Kennewick
7,920,682 B2	4/2011	Byrne	9,269,097 B2	2/2016	Freeman
7,949,529 B2	5/2011	Weider	9,305,548 B2	4/2016	Kennewick
7,949,537 B2	5/2011	Walker	9,308,445 B1	4/2016	Merzenich
7,953,732 B2	5/2011	Frank	9,318,108 B2	4/2016	Gruber
7,974,875 B1	7/2011	Quilici	9,406,078 B2	8/2016	Freeman
7,983,917 B2	7/2011	Kennewick	9,443,514 B1	9/2016	Taubman
7,984,287 B2	7/2011	Gopalakrishnan	9,502,025 B2	11/2016	Kennewick
8,005,683 B2	8/2011	Tessel	9,711,143 B2	7/2017	Kennewick
8,015,006 B2	9/2011	Kennewick	10,089,984 B2	10/2018	Kennewick
8,024,186 B1	9/2011	De Bonet	2001/0011336 A1 *	8/2001	Sitka H04N 1/2179
8,027,965 B2	9/2011	Takehara			711/161
8,032,383 B1	10/2011	Bhardwaj	2001/0014857 A1 *	8/2001	Wang G10L 25/78
8,060,367 B2	11/2011	Keaveney			704/231
8,069,046 B2	11/2011	Kennewick	2001/0021905 A1	9/2001	Burnett
8,073,681 B2	12/2011	Baldwin	2001/0039492 A1	11/2001	Nemoto
8,077,975 B2	12/2011	Ma	2001/0041980 A1	11/2001	Howard
8,082,153 B2	12/2011	Coffman	2001/0047261 A1	11/2001	Kassan
8,086,463 B2	12/2011	Ativanichayaphong	2001/0049601 A1	12/2001	Krocker
8,103,510 B2	1/2012	Sato	2001/0054087 A1	12/2001	Flom
8,112,275 B2	2/2012	Kennewick	2002/0002548 A1	1/2002	Roundtree
8,140,327 B2	3/2012	Kennewick	2002/0007267 A1	1/2002	Batchilo
8,140,335 B2	3/2012	Kennewick	2002/0010584 A1	1/2002	Schultz
8,145,489 B2	3/2012	Freeman	2002/0015500 A1	2/2002	Belt
8,150,694 B2	4/2012	Kennewick	2002/0022927 A1	2/2002	Lemelson
8,155,962 B2	4/2012	Kennewick	2002/0022956 A1	2/2002	Ukrainczyk
8,170,867 B2	5/2012	Germain	2002/0029186 A1	3/2002	Roth
8,180,037 B1	5/2012	Delker	2002/0029261 A1	3/2002	Shibata
8,190,436 B2	5/2012	Bangalore	2002/0032752 A1	3/2002	Gold
8,195,468 B2	6/2012	Weider	2002/0035501 A1	3/2002	Handel
8,200,485 B1	6/2012	Lee	2002/0040297 A1	4/2002	Tsiao
8,204,751 B1	6/2012	Di Fabbri	2002/0049535 A1	4/2002	Rigo
8,219,399 B2	7/2012	Lutz	2002/0049805 A1	4/2002	Yamada
8,219,599 B2	7/2012	Tunstall-Pedoe	2002/0059068 A1	5/2002	Rose
8,224,652 B2	7/2012	Wang	2002/0065568 A1	5/2002	Silfvast
8,255,224 B2	8/2012	Singleton	2002/0067839 A1	6/2002	Heinrich
			2002/0069059 A1	6/2002	Smith
			2002/0069071 A1	6/2002	Knockeart
			2002/0073176 A1	6/2002	Ikeda
			2002/0082911 A1	6/2002	Dunn

US 10,755,699 B2

Page 6

(56)	References Cited		2004/0128514 A1 *	7/2004	Rhoads	B42D 25/29 713/176
	U.S. PATENT DOCUMENTS		2004/0133793 A1	7/2004	Ginter	
			2004/0140989 A1	7/2004	Papageorge	
			2004/0143440 A1	7/2004	Prasad	
2002/0087312 A1	7/2002	Lee	2004/0148154 A1	7/2004	Acero	
2002/0087326 A1	7/2002	Lee	2004/0148170 A1	7/2004	Acero	
2002/0087525 A1	7/2002	Abbott	2004/0158555 A1	8/2004	Seedman	
2002/0107694 A1	8/2002	Lerg	2004/0166832 A1	8/2004	Portman	
2002/0120609 A1	8/2002	Lang	2004/0167771 A1	8/2004	Duan	
2002/0124050 A1	9/2002	Middeljans	2004/0172247 A1	9/2004	Yoon	
2002/0133347 A1	9/2002	Schoneburg	2004/0172258 A1	9/2004	Dominach	
2002/0133354 A1	9/2002	Ross	2004/0189697 A1	9/2004	Fukuoka	
2002/0133402 A1	9/2002	Faber	2004/0193408 A1	9/2004	Hunt	
2002/0135618 A1	9/2002	Maes	2004/0193420 A1	9/2004	Kennewick	
2002/0138248 A1	9/2002	Corston-Oliver	2004/0199375 A1	10/2004	Ehsani	
2002/0143532 A1	10/2002	McLean	2004/0199387 A1	10/2004	Wang	
2002/0143535 A1	10/2002	Kist	2004/0199389 A1	10/2004	Geiger	
2002/0152260 A1	10/2002	Chen	2004/0201607 A1	10/2004	Mulvey	
2002/0161646 A1	10/2002	Gailey	2004/0204939 A1	10/2004	Liu	
2002/0161647 A1	10/2002	Gailey	2004/0205671 A1	10/2004	Sukehiro	
2002/0169597 A1	11/2002	Fain	2004/0220800 A1	11/2004	Kong	
2002/0173333 A1	11/2002	Buchholz	2004/0243393 A1	12/2004	Wang	
2002/0173961 A1	11/2002	Guerra	2004/0243417 A9	12/2004	Pitts	
2002/0178003 A1	11/2002	Gehrke	2004/0247092 A1	12/2004	Timmins	
2002/0184373 A1	12/2002	Maes	2004/0249636 A1	12/2004	Applebaum	
2002/0188455 A1	12/2002	Shioda	2005/0015256 A1 *	1/2005	Kargman	G06Q 30/06 704/272
2002/0188602 A1	12/2002	Stubler	2005/0021331 A1	1/2005	Huang	
2002/0198714 A1	12/2002	Zhou	2005/0021334 A1	1/2005	Iwahashi	
2003/0005033 A1	1/2003	Mohan	2005/0021470 A1	1/2005	Martin	
2003/0014261 A1	1/2003	Kageyama	2005/0021826 A1	1/2005	Kumar	
2003/0016835 A1	1/2003	Elko	2005/0033574 A1	2/2005	Kim	
2003/0036903 A1	2/2003	Konopka	2005/0033582 A1	2/2005	Gadd	
2003/0046071 A1	3/2003	Wyman	2005/0043940 A1	2/2005	Elder	
2003/0046281 A1	3/2003	Son	2005/0043956 A1 *	2/2005	Aoyama	G10L 15/22 704/276
2003/0046346 A1	3/2003	Mumick	2005/0060142 A1	3/2005	Visser	
2003/0064709 A1	4/2003	Gailey	2005/0080632 A1	4/2005	Endo	
2003/0065427 A1	4/2003	Funk	2005/0080821 A1 *	4/2005	Breil	G06Q 40/02
2003/0069734 A1	4/2003	Everhart	2005/0102282 A1	5/2005	Linden	
2003/0069880 A1	4/2003	Harrison	2005/0114116 A1	5/2005	Fiedler	
2003/0088421 A1	5/2003	Maes	2005/0125232 A1	6/2005	Gadd	
2003/0093419 A1	5/2003	Bangalore	2005/0131673 A1	6/2005	Koizumi	
2003/0097249 A1	5/2003	Walker	2005/0137850 A1	6/2005	Odell	
2003/0110037 A1	6/2003	Walker	2005/0137877 A1	6/2005	Oesterling	
2003/0112267 A1	6/2003	Belrose	2005/0138168 A1	6/2005	Hoffman	
2003/0115062 A1	6/2003	Walker	2005/0143994 A1	6/2005	Mori	
2003/0120493 A1	6/2003	Gupta	2005/0144013 A1	6/2005	Fujimoto	
2003/0135488 A1	7/2003	Amir	2005/0144187 A1	6/2005	Che	
2003/0144846 A1	7/2003	Denenberg	2005/0149319 A1	7/2005	Honda	
2003/0154082 A1 *	8/2003	Toguri	2005/0169441 A1 *	8/2005	Yacoub	H04M 3/4936 379/88.16
		G10L 17/00 704/268	2005/0216254 A1	9/2005	Gupta	
2003/0158731 A1	8/2003	Falcon	2005/0222763 A1	10/2005	Uyeki	
2003/0161448 A1	8/2003	Parolkar	2005/0234637 A1	10/2005	Obradovich	
2003/0167167 A1 *	9/2003	Gong	2005/0234727 A1	10/2005	Chiu	
		G10L 15/22 704/250	2005/0240412 A1 *	10/2005	Fujita	G10L 15/16 704/270
2003/0174155 A1	9/2003	Weng	2005/0246174 A1	11/2005	DeGolia	
2003/0182132 A1	9/2003	Niemoeller	2005/0283364 A1	12/2005	Longe	
2003/0187643 A1	10/2003	VanThong	2005/0283532 A1	12/2005	Kim	
2003/0204492 A1	10/2003	Wolf	2005/0283752 A1	12/2005	Fruchter	
2003/0206640 A1	11/2003	Malvar	2006/0041431 A1	2/2006	Maes	
2003/0212550 A1	11/2003	Ubale	2006/0046740 A1	3/2006	Johnson	
2003/0212558 A1	11/2003	Marula	2006/0047509 A1	3/2006	Ding	
2003/0212562 A1	11/2003	Patel	2006/0072738 A1	4/2006	Louis	
2003/0225825 A1	12/2003	Healey	2006/0074670 A1	4/2006	Weng	
2003/0233230 A1	12/2003	Ammicht	2006/0074671 A1	4/2006	Farmaner	
2003/0236664 A1	12/2003	Sharma	2006/0080098 A1	4/2006	Campbell	
2004/0006475 A1	1/2004	Ehlen	2006/0100851 A1	5/2006	Schonebeck	
2004/0010358 A1	1/2004	Oesterling	2006/0106769 A1	5/2006	Gibbs	
2004/0025115 A1	2/2004	Sienel	2006/0129409 A1	6/2006	Mizutani	
2004/0030741 A1	2/2004	Wotton	2006/0130002 A1	6/2006	Hirayama	
2004/0036601 A1	2/2004	Obradovich	2006/0149633 A1	7/2006	Voisin	
2004/0044516 A1 *	3/2004	Kennewick	2006/0182085 A1	8/2006	Sweeney	
		G10L 15/22 704/5	2006/0206310 A1	9/2006	Ravikumar	
2004/0093567 A1	5/2004	Schabes	2006/0217133 A1	9/2006	Christenson	
2004/0098245 A1	5/2004	Walker	2006/0236343 A1	10/2006	Chang	
2004/0102977 A1	5/2004	Metzler				
2004/0117179 A1	6/2004	Balasuriya				
2004/0117804 A1	6/2004	Scahill				
2004/0122673 A1	6/2004	Park				
2004/0122674 A1	6/2004	Bangalore				

US 10,755,699 B2

Page 7

(56)

References Cited

U.S. PATENT DOCUMENTS

2006/0242017	A1	10/2006	Libes		2008/0162471	A1	7/2008	Bernard
2006/0247919	A1*	11/2006	Specht	 H04K 3/825	2008/0177530	A1	7/2008	Cross
				704/201	2008/0184164	A1	7/2008	Di Fabbizio
2006/0253247	A1	11/2006	de Silva		2008/0189110	A1	8/2008	Freeman
2006/0253281	A1	11/2006	Letzt		2008/0189187	A1	8/2008	Hao
2006/0285662	A1	12/2006	Yin		2008/0228496	A1	9/2008	Yu
2007/0011159	A1	1/2007	Hillis		2008/0235023	A1	9/2008	Kennewick
2007/0033005	A1	2/2007	Di Cristo		2008/0235027	A1	9/2008	Cross
2007/0033020	A1	2/2007	Francois		2008/0269958	A1	10/2008	Filev
2007/0033526	A1	2/2007	Thompson		2008/0270135	A1	10/2008	Goel
2007/0038436	A1	2/2007	Cristo		2008/0270224	A1	10/2008	Portman
2007/0038445	A1	2/2007	Helbing		2008/0294437	A1	11/2008	Nakano
2007/0043569	A1	2/2007	Potter		2008/0294994	A1	11/2008	Kruger
2007/0043574	A1	2/2007	Coffman		2008/0306743	A1	12/2008	Di Fabbizio
2007/0043868	A1	2/2007	Kumar		2008/0319751	A1	12/2008	Kennewick
2007/0050191	A1	3/2007	Weider		2009/0006077	A1	1/2009	Keaveney
2007/0050279	A1	3/2007	Huang		2009/0006194	A1	1/2009	Sridharan
2007/0055525	A1	3/2007	Kennewick		2009/0018829	A1	1/2009	Kuperstein
2007/0060114	A1	3/2007	Ramer		2009/0024476	A1	1/2009	Baar
2007/0061067	A1	3/2007	Zeinstra		2009/0030686	A1	1/2009	Weng
2007/0061735	A1	3/2007	Hoffberg		2009/0052635	A1	2/2009	Jones
2007/0067310	A1	3/2007	Gupta		2009/0055176	A1	2/2009	Hu
2007/0073544	A1	3/2007	Millett		2009/0067599	A1	3/2009	Agarwal
2007/0078708	A1	4/2007	Yu		2009/0076827	A1	3/2009	Bulitta
2007/0078709	A1	4/2007	Rajaram		2009/0106029	A1	4/2009	DeLine
2007/0078814	A1	4/2007	Flowers		2009/0117885	A1	5/2009	Roth
2007/0094003	A1	4/2007	Huang		2009/0144131	A1	6/2009	Chiu
2007/0100797	A1	5/2007	Thun		2009/0144271	A1	6/2009	Richardson
2007/0106499	A1	5/2007	Dahlgren		2009/0150156	A1	6/2009	Kennewick
2007/0112555	A1	5/2007	Lavi		2009/0157382	A1	6/2009	Bar
2007/0112630	A1	5/2007	Lau		2009/0164216	A1	6/2009	Chengalvarayan
2007/0118357	A1	5/2007	Kasravi		2009/0171664	A1	7/2009	Kennewick
2007/0124057	A1	5/2007	Prieto		2009/0171912	A1	7/2009	Nash
2007/0135101	A1	6/2007	Ramati		2009/0197582	A1	8/2009	Lewis
2007/0146833	A1	6/2007	Satomi		2009/0216540	A1	8/2009	Tessel
2007/0162296	A1	7/2007	Altberg		2009/0248565	A1	10/2009	Chuang
2007/0174258	A1	7/2007	Jones		2009/0248605	A1	10/2009	Mitchell
2007/0179778	A1	8/2007	Gong		2009/0259561	A1	10/2009	Boys
2007/0185859	A1	8/2007	Flowers		2009/0259646	A1	10/2009	Fujita
2007/0186165	A1	8/2007	Maislos		2009/0265163	A1	10/2009	Li
2007/0192309	A1	8/2007	Fischer		2009/0271194	A1	10/2009	Davis
2007/0198267	A1	8/2007	Jones		2009/0273563	A1	11/2009	Pryor
2007/0203699	A1	8/2007	Nagashima		2009/0276700	A1	11/2009	Anderson
2007/0203736	A1	8/2007	Ashton		2009/0287680	A1	11/2009	Paek
2007/0208732	A1	9/2007	Flowers		2009/0299745	A1	12/2009	Kennewick
2007/0214182	A1	9/2007	Rosenberg		2009/0299857	A1	12/2009	Brubaker
2007/0250901	A1	10/2007	McIntire		2009/0304161	A1	12/2009	Pettyjohn
2007/0265850	A1	11/2007	Kennewick		2009/0307031	A1	12/2009	Winkler
2007/0266257	A1	11/2007	Camaisa		2009/0313026	A1	12/2009	Coffman
2007/0276651	A1	11/2007	Bliss		2009/0319517	A1	12/2009	Guha
2007/0294615	A1	12/2007	Sathe		2010/0023320	A1	1/2010	Cristo
2007/0299824	A1	12/2007	Pan		2010/0023331	A1	1/2010	Duta
2008/0010135	A1	1/2008	Schrock		2010/0029261	A1	2/2010	Mikkelsen
2008/0014908	A1	1/2008	Vasant		2010/0036967	A1	2/2010	Caine
2008/0034032	A1	2/2008	Healey		2010/0049501	A1	2/2010	Kennewick
2008/0046311	A1	2/2008	Shahine		2010/0049514	A1	2/2010	Kennewick
2008/0059188	A1	3/2008	Konopka		2010/0057443	A1	3/2010	Cristo
2008/0065386	A1	3/2008	Cross		2010/0063880	A1	3/2010	Atsmon
2008/0065389	A1	3/2008	Cross		2010/0064025	A1	3/2010	Nelimarkka
2008/0065390	A1	3/2008	Ativanichayaphong		2010/0076778	A1	3/2010	Kondrk
2008/0086455	A1	4/2008	Meisels		2010/0094707	A1	4/2010	Freer
2008/0091406	A1	4/2008	Baldwin		2010/0138300	A1	6/2010	Wallis
2008/0103761	A1	5/2008	Printz		2010/0145700	A1	6/2010	Kennewick
2008/0103781	A1	5/2008	Wasson		2010/0185512	A1	7/2010	Borger
2008/0104071	A1	5/2008	Pragada		2010/0191856	A1	7/2010	Gupta
2008/0109285	A1	5/2008	Reuther		2010/0204986	A1	8/2010	Kennewick
2008/0115163	A1	5/2008	Gilboa		2010/0204994	A1	8/2010	Kennewick
2008/0126091	A1	5/2008	Clark		2010/0217604	A1	8/2010	Baldwin
2008/0126284	A1	5/2008	Forbes		2010/0268536	A1	10/2010	Suendermann
2008/0133215	A1	6/2008	Sarukkai		2010/0286985	A1	11/2010	Kennewick
2008/0140385	A1	6/2008	Mahajan		2010/0299142	A1	11/2010	Freeman
2008/0147396	A1	6/2008	Wang		2010/0312547	A1	12/2010	Van Os
2008/0147410	A1	6/2008	Odinak		2010/0312566	A1	12/2010	Odinak
2008/0147637	A1	6/2008	Li		2010/0318357	A1	12/2010	Istvan
2008/0154604	A1	6/2008	Sathish		2010/0331064	A1	12/2010	Michelstein
					2011/0022393	A1	1/2011	Waller
					2011/0106527	A1	5/2011	Chiu
					2011/0112827	A1	5/2011	Kennewick
					2011/0112921	A1	5/2011	Kennewick

US 10,755,699 B2

Page 8

(56)

References Cited

U.S. PATENT DOCUMENTS

2011/0119049	A1	5/2011	Ylonen
2011/0131036	A1	6/2011	DiCristo
2011/0131045	A1	6/2011	Cristo
2011/0231182	A1	9/2011	Weider
2011/0231188	A1	9/2011	Kennewick
2011/0238409	A1	9/2011	Larcheveque
2011/0307167	A1	12/2011	Taschereau
2012/0022857	A1	1/2012	Baldwin
2012/0041753	A1	2/2012	Dymetman
2012/0046935	A1	2/2012	Nagao
2012/0101809	A1	4/2012	Kennewick
2012/0101810	A1	4/2012	Kennewick
2012/0109753	A1	5/2012	Kennewick
2012/0150620	A1	6/2012	Mandyam
2012/0150636	A1	6/2012	Freeman
2012/0239498	A1	9/2012	Ramer
2012/0240060	A1	9/2012	Pennington
2012/0265528	A1	10/2012	Gruber
2012/0278073	A1	11/2012	Weider
2013/0006734	A1	1/2013	Ocko
2013/0054228	A1	2/2013	Baldwin
2013/0060625	A1	3/2013	Davis
2013/0080177	A1	3/2013	Chen
2013/0211710	A1	8/2013	Kennewick
2013/0253929	A1	9/2013	Weider
2013/0254314	A1	9/2013	Chow
2013/0297293	A1	11/2013	Cristo
2013/0304473	A1	11/2013	Baldwin
2013/0311324	A1	11/2013	Stoll
2013/0332454	A1	12/2013	Stuhec
2013/0339022	A1	12/2013	Baldwin
2014/0006951	A1	1/2014	Hunter
2014/0012577	A1	1/2014	Freeman
2014/0025371	A1	1/2014	Min
2014/0108013	A1	4/2014	Cristo
2014/0136187	A1	5/2014	Wolverton
2014/0156278	A1	6/2014	Kennewick
2014/0195238	A1	7/2014	Terao
2014/0236575	A1	8/2014	Tur
2014/0249821	A1	9/2014	Kennewick
2014/0249822	A1	9/2014	Baldwin
2014/0278413	A1	9/2014	Pitschel
2014/0278416	A1	9/2014	Schuster
2014/0288934	A1	9/2014	Kennewick
2014/0330552	A1	11/2014	Bangalore
2014/0337007	A1	11/2014	Waibel
2014/0365222	A1	12/2014	Weider
2015/0019211	A1	1/2015	Simard
2015/0019217	A1	1/2015	Cristo
2015/0019227	A1	1/2015	Anandarajah
2015/0066479	A1	3/2015	Pasupalak
2015/0066627	A1	3/2015	Freeman
2015/0073910	A1	3/2015	Kennewick
2015/0095159	A1	4/2015	Kennewick
2015/0142447	A1	5/2015	Kennewick
2015/0170641	A1	6/2015	Kennewick
2015/0193379	A1	7/2015	Mehta
2015/0199339	A1	7/2015	Mirkin
2015/0228276	A1	8/2015	Baldwin
2015/0293917	A1	10/2015	Bufe
2015/0348544	A1	12/2015	Baldwin
2015/0348551	A1	12/2015	Gruber
2015/0364133	A1	12/2015	Freeman
2016/0049152	A1	2/2016	Kennewick
2016/0078482	A1	3/2016	Kennewick
2016/0078491	A1	3/2016	Kennewick
2016/0078504	A1	3/2016	Kennewick
2016/0078773	A1	3/2016	Carter
2016/0110347	A1	4/2016	Kennewick
2016/0148610	A1	5/2016	Kennewick
2016/0148612	A1	5/2016	Guo
2016/0188292	A1	6/2016	Carter
2016/0188573	A1	6/2016	Tang

2016/0217785	A1	7/2016	Kennewick
2016/0335676	A1	11/2016	Freeman
2017/0004588	A1	1/2017	Isaacson

FOREIGN PATENT DOCUMENTS

CN	101236635	A	8/2008
CN	103229206	A	7/2013
EP	1320043	A2	6/2003
EP	1646037		4/2006
FR	2867583		9/2005
JP	H08263258		10/1996
JP	H11249773		9/1999
JP	2001071289		3/2001
JP	2006146881		6/2006
JP	2008027454		2/2008
JP	2008058465		3/2008
JP	2008139928		6/2008
JP	2011504304		2/2011
JP	2012518847		8/2012
WO	9946763		9/1999
WO	0021232		4/2000
WO	0046792		8/2000
WO	0129742		4/2001
WO	0129742	A2	4/2001
WO	0171609	A2	9/2001
WO	0178065		10/2001
WO	2004072954		8/2004
WO	2005010702	A2	2/2005
WO	2007019318		1/2007
WO	2007021587		1/2007
WO	2007027546		1/2007
WO	2007027989		1/2007
WO	2008098039		1/2008
WO	2008118195		1/2008
WO	2009075912		1/2009
WO	2009145796		1/2009
WO	2009111721		9/2009
WO	2010096752		1/2010
WO	2016044290		3/2016
WO	2016044316		3/2016
WO	2016044319		3/2016
WO	2016044321		3/2016
WO	2016061309		4/2016

OTHER PUBLICATIONS

Arrington, Michael, "Google Redefines GPS Navigation Landscape: Google Maps Navigation for Android 2.0", TechCrunch, printed from the Internet <<http://www.techcrunch.com/2009/10/28/google-redetines-car-gps-navigation-google-maps-navigation-android/>>, Oct. 28, 2009, 4 pages.

Bazzi, Issam et al., "Heterogeneous Lexical Units for Automatic Speech Recognition: Preliminary Investigations", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, Jun. 5-9, 2000, XP010507574, pp. 1257-1260.

Belvin, Robert, et al., "Development of the HRL Route Navigation Dialogue System", Proceedings of the First International Conference on Human Language Technology Research, San Diego, 2001, pp. 1-5.

Chai et al., "MIND: A Semantics-Based Multimodal Interpretation Framework for Conversational Systems", Proceedings of the International Class Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Jun. 2002, pp. 37-46.

Cheyen et al., "Multimodal Maps: An Agent-Based Approach", International Conference on Cooperative Multimodal Communication (CMC/95), May 24-26, 1995, pp. 111-121.

Davis, Z., et al., A Personal Handheld Multi-Modal Shopping Assistant, IEEE, 2006, 9 pages.

El Meliani et al., "A Syllabic-Filler-Based Continuous Speech Recognizer for Unlimited Vocabulary", Canadian Conference on Electrical and Computer Engineering, vol. 2, Sep. 5-8, 1995, pp. 1007-1010.

Elio et al., "On Abstract Task Models and Conversation Policies" in Workshop on Specifying and Implementing Conversation Policies, Autonomous Agents '99, Seattle, 1999, 10 pages.

US 10,755,699 B2

Page 9

(56)

References Cited

OTHER PUBLICATIONS

Kirchhoff, Katrin, "Syllable-Level Desynchronisation of Phonetic Features for Speech Recognition", Proceedings of the Fourth International Conference on Spoken Language, 1996, ICSLP 96, vol. 4, IEEE, 1996, 3 pages.

Kuhn, Thomas, et al., "Hybrid In-Car Speech Recognition for Mobile Multimedia Applications", Vehicular Technology Conference, IEEE, Jul. 1999, pp. 2009-2013.

Lin, Bor-shen, et al., "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History", ASRU'99, 1999, 4 pages.

Lind, R., et al., "The Network Vehicle—A Glimpse into the Future of Mobile Multi-Media", IEEE Aerosp. Electron. Systems Magazine, vol. 14, No. 9, Sep. 1999, pp. 27-32.

Mao, Mark Z., "Automatic Training Set Segmentation for Multi-Pass Speech Recognition", Department of Electrical Engineering, Stanford University, CA, copyright 2005, IEEE, pp. 1-685 to 1-688.

O'Shaughnessy, Douglas, "Interacting with Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, vol. 91, No. 9, Sep. 1, 2003, XP011100665, pp. 1272-1305.

Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., Jul. 28, 2002, 1 page.

Turunen, "Adaptive Interaction Methods in Speech User Interfaces", Conference on Human Factors in Computing Systems, Seattle, Washington, 2001, pp. 91-92.

Vanhoutte, Vincent, "Confidence Scoring and Rejection Using Multi-Pass Speech Recognition", Nuance Communications, Menlo Park, CA, 2005, 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, 1998, 4 pages.

Wu, Su-Lin, et al., "Incorporating Information from Syllable-Length Time Scales into Automatic Speech Recognition", Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, vol. 2 IEEE, 1998, 4 pages.

Wu, Su-Lin, et al., "Integrating Syllable Boundary Information into Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, 1997, vol. 2, IEEE, 1997, 4 pages.

Zhao, Yilin, "Telematics: Safe and Fun Driving", IEEE Intelligent Systems, vol. 17, Issue 1, 2002, pp. 10-14.

Office Action issued in Chinese Patent Application No. 201580060519.1 dated Mar. 13, 2020, with its English translation, 40 pages.

* cited by examiner

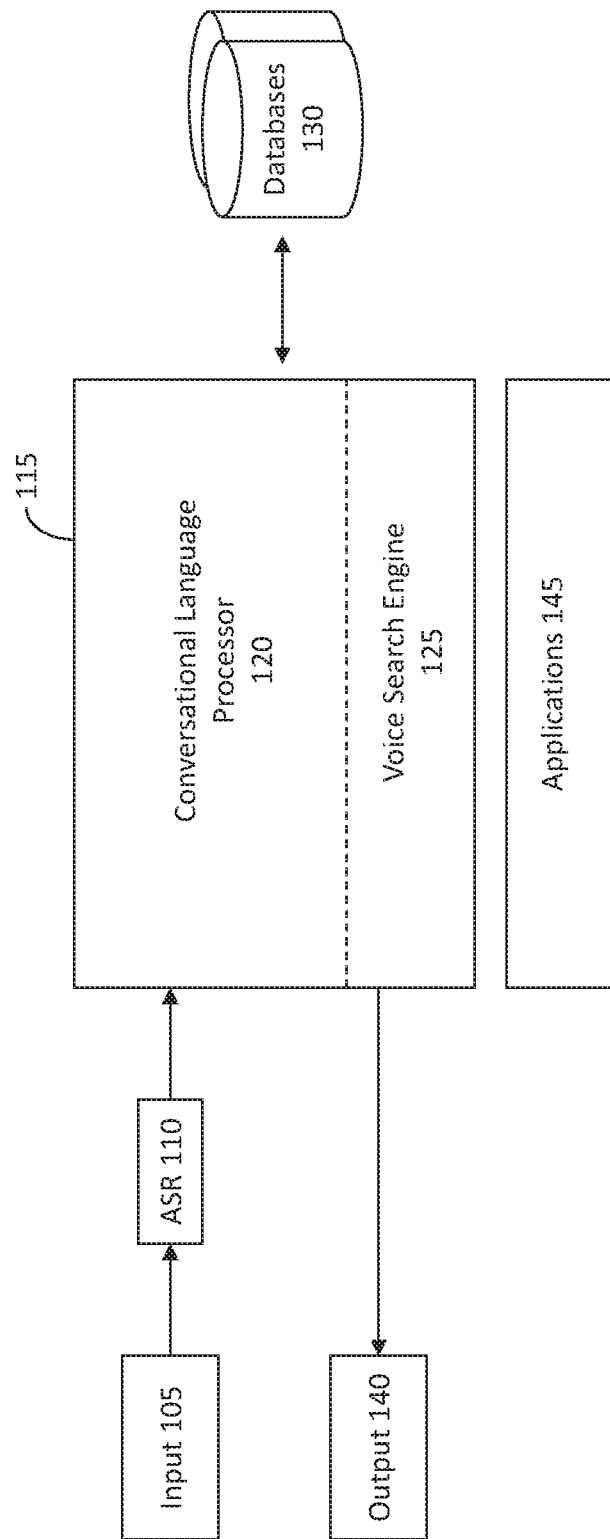


Figure 1

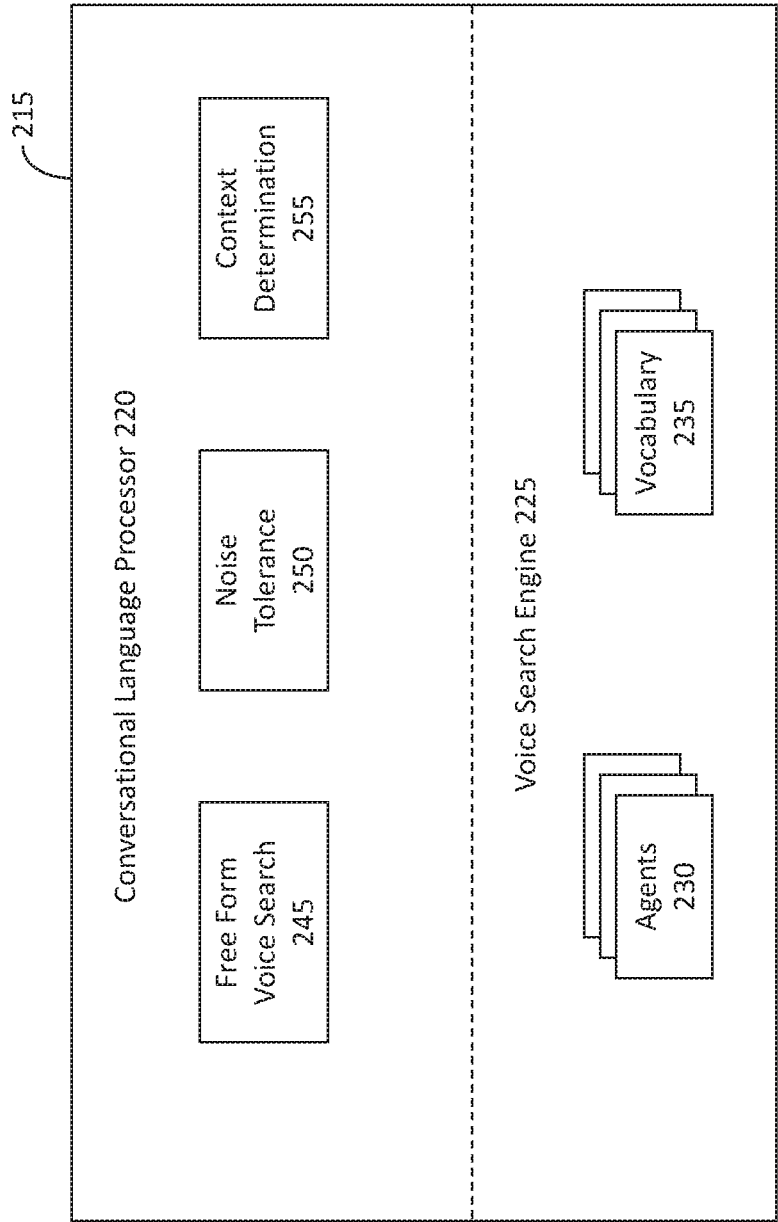


Figure 2

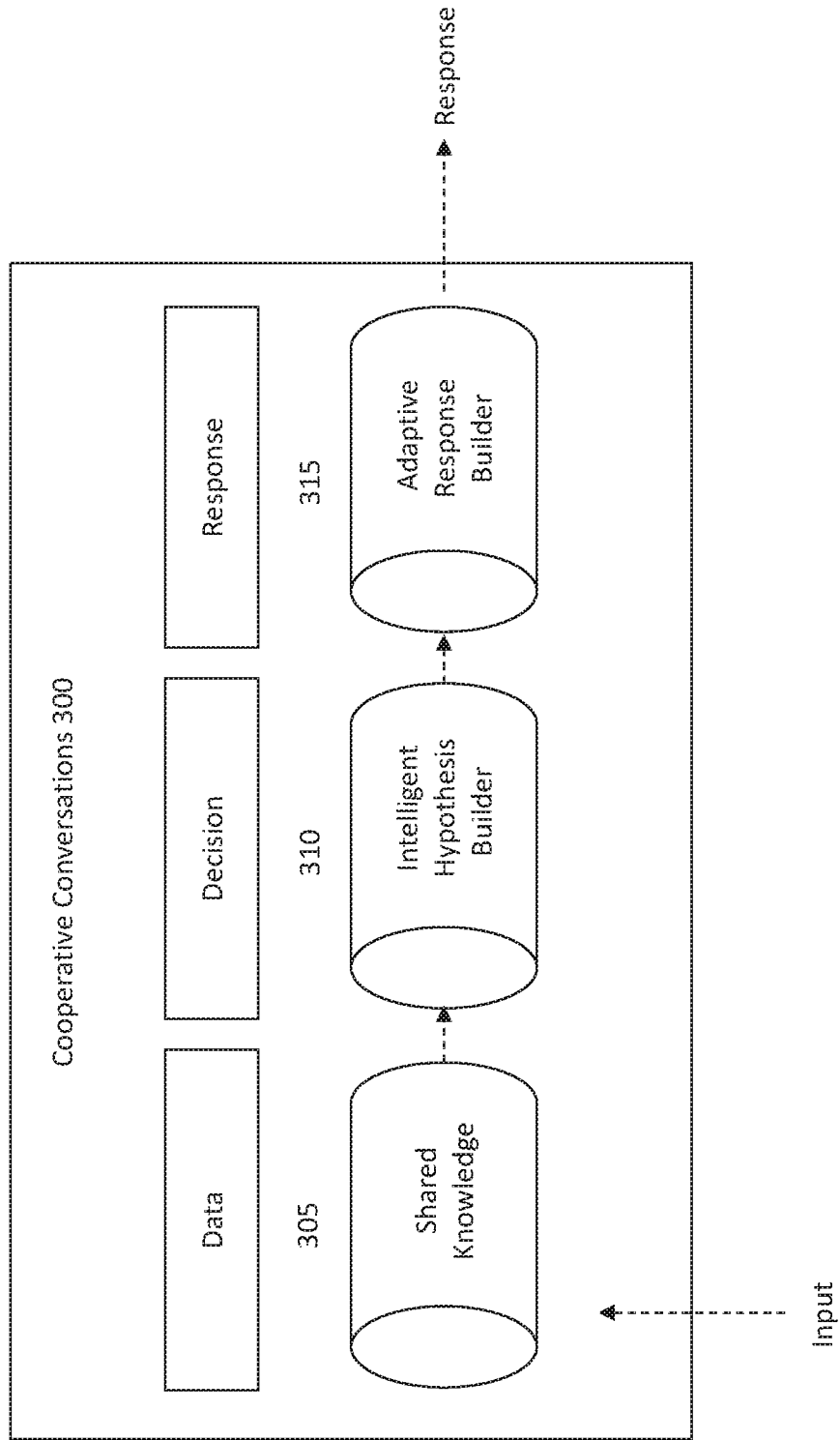


Figure 3

US 10,755,699 B2

1

SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a divisional of U.S. patent application Ser. No. 14/691,445, entitled "SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE," filed Apr. 20, 2015, which is a continuation of U.S. patent application Ser. No. 13/987,645, entitled "SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE," filed Aug. 19, 2013 (which issued as U.S. Pat. No. 9,015,049 on Apr. 21, 2015), which is a divisional of U.S. patent application Ser. No. 13/251,712, entitled "SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE," filed Oct. 3, 2011 (which issued as U.S. Pat. No. 8,515,765 on Aug. 20, 2013), which is a continuation of U.S. patent application Ser. No. 11/580,926, entitled "SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE," filed Oct. 16, 2006 (which issued as U.S. Pat. No. 8,073,681 on Dec. 6, 2011), each of which are hereby incorporated by reference in its entirety.

FIELD OF THE INVENTION

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

BACKGROUND OF THE INVENTION

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.

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

2

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.

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

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.

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 a speech recognition engine (which may alternatively be referred to as an Automatic Speech Recognizer or ASR) 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.

According to another aspect of the invention, an exemplary conversational 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

US 10,755,699 B2

3

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

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.

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

4

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 agent “wins” the competition, the winning domain 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. 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 out to be incorrect, the incorrect interpretation may be removed as a potential interpretation from one or more grammars associated with the speech recognition engine and/or from possible interpretations determined by the conversational speech engine, thereby assuring that a mistake will not be repeated for an identical utterance.

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.

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 conversational model 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. By taking advantage of human presumptions about utterances that humans rely upon, both

US 10,755,699 B2

5

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).

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 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.

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

6

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).

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.

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 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).

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."

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.

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

US 10,755,699 B2

7

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.

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 be chosen in subsequent responses to move conversation forward and establish an appropriate domain to be explored with the user.

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

FIG. 1 is an exemplary block diagram of a system architecture according to one aspect of the invention.

FIG. 2 is an exemplary block diagram of a conversational speech engine according to one aspect of the invention.

FIG. 3 is an exemplary block diagram of a cooperative conversational model according to one aspect of the invention.

DETAILED DESCRIPTION

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

The utterance component of input **105** may be processed by a speech recognition engine **110** (which may alternatively be referred to herein as Automatic Speech Recognizer **110**, or as shown in FIG. 1, ASR **110**) to generate one or more preliminary interpretations of the utterance. The speech recognition engine **110** may process the utterance using any suitable technique known in the art. For example, in one implementation, the speech recognition engine **110** may interpret the utterance using techniques of phonetic dictation to recognize a phoneme stream, as described in U.S. patent application Ser. No. 11/513,269, entitled “Dynamic Speech Sharpening,” filed Aug. 31, 2006, which issued as U.S. Pat. No. 7,634,409 on Dec. 15, 2009, and which is hereby incorporated by reference in its entirety. The one or more preliminary interpretations generated by the speech recognition engine **110** may then be provided to a conversational speech engine **115** for further processing. Conversational speech engine **115** may include a conversational language

8

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, the speech recognition engine **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).

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 domain agents **230** and/or one or more vocabularies **235** to aid in interpreting utterances and generating responses, as described in “Enhancing the VUE™ (Voice-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.

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.

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.,

US 10,755,699 B2

9

“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 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).

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.

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. 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

10

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.

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.

The imperfect speech feature may be able to infer requests 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, 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.

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,

US 10,755,699 B2

11

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 those conditions.

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 U.S. patent application Ser. No. 11/197,504, entitled “Systems and Methods for Responding to Natural Language Speech Utterance,” filed Aug. 5, 2005, which issued as U.S. Pat. No. 7,640,160 on Dec. 29, 2009 and U.S. patent application Ser. No. 11/212,693, entitled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions,” filed Aug. 29, 2005, which issued as U.S. Pat. No. 7,949,529 on May 24, 2011, 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).

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 grammars associated with the speech recognition engine 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.

12

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

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, Ore. based on the Weather and an environmental proximity between Portland, Ore. and Seattle, Wash. 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, N.H.,” in the second utterance, context determination process **255** may instead retrieve a weather report for Portland, Me. 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 disambiguate between possibilities, responses may prompt the user with a request for further information (e.g., “Did you mean Portland, Me., or Portland, Ore.?”).

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

US 10,755,699 B2

13

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.

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 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.

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

14

building) **310**, and then may refine the decision and generate a response (adaptive response building) **315**.

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.

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

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 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.

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,

US 10,755,699 B2

15

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.

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).

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

16

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

EXPLORATORY		
	Participant A User	Participant B 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

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

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.

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 utter-

US 10,755,699 B2

17

ance. 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.

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, 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.

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 a confidence level associated with a preliminary interpretation generated at the speech recognition engine 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 confidence level associated with the preliminary interpretation generated at the speech recognition engine is below the predetermined threshold; (3) "not sure," where additional contextual signifiers or qualifiers are

18

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 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).

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.

According to another aspect of the invention, Adaptive Response Builder 310 may draw on information maintained by Intelligent 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.

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 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?

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?
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?

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.

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.

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 be 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:

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.

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.

What is claimed is:

1. A computer-implemented method of generating natural language system responses adapted based on a user’s manner of speaking, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:
- receiving, by the computer system, a user input comprising a natural language utterance;
 - recognizing, by the computer system, one or more words or phrases from the natural language utterance;
 - identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;
 - determining, by the computer system, an interpretation of the natural language utterance based on the identified context;
 - accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;
 - accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;
 - identifying, by the computer system, a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and

US 10,755,699 B2

21

generating, by the computer system, a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.

2. The method of claim 1, wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.

3. The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response based on the identified manner comprises varying one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.

4. The method of claim 1, the method further comprising: obtaining, by the computer system, contextual signifiers and/or grammatical rules, wherein generating the response based on the identified manner in which the natural language utterance was spoken comprises using the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.

5. The method of claim 1, wherein the long-term knowledge is associated with a first user, the method further comprising:

generating, by the computer system, a profile associated with the first user based on the long-term knowledge, wherein the context for the natural language utterance is determined based further on the profile associated with the first user.

6. The method of claim 1, wherein generating the response based on the identified manner comprises:

adapting, by the computer system, the response based on a response format associated with the identified manner.

7. The method of claim 1, wherein recognizing the one or more words or phrases from the natural language utterance comprises:

providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and

obtaining, by the computer system, the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.

8. The method of claim 1, the method further comprising: causing, by the computer system, the response to the natural language utterance to be provided to the user.

9. The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response to the natural language utterance based on the identified manner comprises varying one or more of word use and/or jargon of the voice response.

10. The method of claim 1, the method further comprising:

adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.

11. The method of claim 1, the method further comprising generating a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.

12. A system for generating natural language system responses adapted based on a user's manner of speaking, the system comprising:

22

one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to:

receive a user input comprising a natural language utterance;

recognize one or more words or phrases from the natural language utterance;

identify a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;

determine an interpretation of the natural language utterance based on the identified context;

accumulate short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system; accumulate long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;

identify a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and

generate a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.

13. The system of claim 12, wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.

14. The system of claim 12, wherein the response comprises a voice response, and wherein to generate the voice response based on the identified manner, the one or more physical processors are further configured to:

vary one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.

15. The system of claim 12, wherein the one or more physical processors are further configured to:

obtain contextual signifiers and/or grammatical rules, wherein to generate the response based on the identified manner in which the natural language utterance was spoken, the one or more physical processors are further configured to use the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.

16. The system of claim 12, wherein the long-term knowledge is associated with a first user, and wherein the one or more physical processors are further configured to: generate a profile associated with the first user based on the long-term knowledge, wherein the context for the natural language utterance is determined based further on the profile associated with the first user.

17. The system of claim 12, wherein to generate the response based on the identified manner, the one or more physical processors are configured to:

adapt the response based on a response format associated with the identified manner.

18. The system of claim 12, wherein to recognize the one or more words or phrases from the natural language utterance, the one or more physical processors are configured to:

provide the natural language utterance as an input to a speech recognition engine; and

provide the natural language utterance as an input to a speech recognition engine; and

US 10,755,699 B2

23

24

obtain the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.

19. The system of claim 12, wherein the one or more physical processors are further configured to: 5
cause the response to the natural language utterance to be provided to the user.

20. The system of claim 12, wherein the response comprises a voice response, and wherein to generate the voice response to the natural language utterance based on the identified manner, the one or more physical processors are further configured to: 10

vary one or more of word use and/or jargon of the voice response.

21. The system of claim 12, wherein the one or more physical processors are further configured to: 15

adapt the response to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use. 20

22. The system of claim 12, wherein the one or more physical processors are further configured to generate a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances. 25

* * * * *

EXHIBIT C



US011087385B2

(12) **United States Patent**
Kennewick, Sr.

(10) **Patent No.:** **US 11,087,385 B2**

(45) **Date of Patent:** **Aug. 10, 2021**

(54) **VOICE COMMERCE**

(56) **References Cited**

(71) Applicant: **VB Assets, LLC**, Bellevue, WA (US)

U.S. PATENT DOCUMENTS

(72) Inventor: **Michael R. Kennewick, Sr.**, Bellevue, WA (US)

4,430,669 A 2/1984 Cheung
4,821,027 A 4/1989 Mallory
(Continued)

(73) Assignee: **VB Assets, LLC**, Bellevue, WA (US)

FOREIGN PATENT DOCUMENTS

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

CN 1433554 A 7/2003
CN 1860496 A 11/2006
(Continued)

(21) Appl. No.: **16/553,553**

OTHER PUBLICATIONS

(22) Filed: **Aug. 28, 2019**

Office Action issued in Chinese Patent Application No. 201580060519.1 dated Mar. 13, 2020, with its English translation, 40 pages.

(65) **Prior Publication Data**

US 2019/0385218 A1 Dec. 19, 2019

(Continued)

Primary Examiner — Mila Airapetian

(74) *Attorney, Agent, or Firm* — Sheppard Mullin Richter & Hampton LLP

Related U.S. Application Data

(63) Continuation of application No. 15/484,930, filed on Apr. 11, 2017, now Pat. No. 10,430,863, which is a (Continued)

(51) **Int. Cl.**

G06Q 30/00 (2012.01)

G06Q 30/06 (2012.01)

(Continued)

(52) **U.S. Cl.**

CPC **G06Q 30/0635** (2013.01); **G06Q 30/0601** (2013.01); **G06Q 30/0625** (2013.01);

(Continued)

(58) **Field of Classification Search**

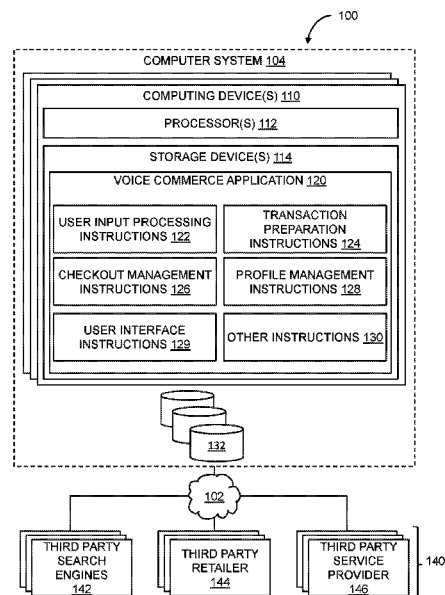
None

See application file for complete search history.

(57) **ABSTRACT**

In certain implementations, a system for facilitating voice commerce is provided. A user input comprising a natural language utterance related to a product or service to be purchased may be received. A first product or service that is to be purchased may be determined based on the utterance. First payment information that is to be used to purchase the first product or service may be obtained. First shipping information that is to be used to deliver the first product or service may be obtained. A purchase transaction for the first product or service may be completed based on the first payment information and the first shipping information without further user input, after the receipt of utterance, that identifies a product or service type or a product or service, seller information, payment information, shipping information, or other information related to purchasing the first product or service.

40 Claims, 6 Drawing Sheets



US 11,087,385 B2

Page 2

Related U.S. Application Data

continuation of application No. 14/855,334, filed on
Sep. 15, 2015, now Pat. No. 9,626,703.

(60) Provisional application No. 62/051,273, filed on Sep.
16, 2014.

(51) **Int. Cl.**

G10L 15/22 (2006.01)

G10L 15/18 (2013.01)

(52) **U.S. Cl.**

CPC **G10L 15/1815** (2013.01); **G10L 15/22**
(2013.01); **G10L 15/18** (2013.01); **G10L**
2015/223 (2013.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,829,423 A	5/1989	Tennant	5,837,721 A	11/1998	Jones
4,887,212 A	12/1989	Zamora	5,839,107 A	11/1998	Gupta
4,910,784 A	3/1990	Doddington	5,848,396 A	12/1998	Gerace
5,027,406 A	6/1991	Roberts	5,855,000 A	12/1998	Waibel
5,155,743 A	10/1992	Jacobs	5,860,059 A	1/1999	Aust
5,164,904 A	11/1992	Sumner	5,867,817 A	2/1999	Catallo
5,208,748 A	5/1993	Flores	5,878,385 A	3/1999	Bralich
5,265,065 A	11/1993	Turtle	5,878,386 A	3/1999	Coughlin
5,274,560 A	12/1993	LaRue	5,892,813 A	4/1999	Morin
5,331,554 A	7/1994	Graham	5,892,900 A	4/1999	Ginter
5,357,596 A	10/1994	Takebayashi	5,895,464 A	4/1999	Bhandari
5,369,575 A	11/1994	Lamberti	5,895,466 A	4/1999	Goldberg
5,377,350 A	12/1994	Skinner	5,897,613 A	4/1999	Chan
5,386,556 A	1/1995	Hedin	5,898,760 A	4/1999	Smets
5,424,947 A	6/1995	Nagao	5,899,991 A	5/1999	Karch
5,471,318 A	11/1995	Ahuja	5,902,347 A	5/1999	Backman
5,475,733 A	12/1995	Eisdorter	5,911,120 A	6/1999	Jarett
5,479,563 A	12/1995	Yamaguchi	5,918,222 A	6/1999	Fukui
5,488,652 A	1/1996	Bielby	5,926,784 A	7/1999	Richardson
5,499,289 A	3/1996	Bruno	5,933,822 A	8/1999	Braden-Harder
5,500,920 A	3/1996	Kupiec	5,950,167 A	9/1999	Yaker
5,517,560 A	5/1996	Greenspan	5,953,393 A	9/1999	Culbreth
5,533,108 A	7/1996	Harris	5,960,384 A	9/1999	Brash
5,537,436 A	7/1996	Bottoms	5,960,397 A	9/1999	Rahim
5,539,744 A	7/1996	Chu	5,960,399 A	9/1999	Barclay
5,557,667 A	9/1996	Bruno	5,960,447 A	9/1999	Holt
5,559,864 A	9/1996	Kennedy, Jr.	5,963,894 A	10/1999	Richardson
5,563,937 A	10/1996	Bruno	5,963,940 A	10/1999	Liddy
5,577,165 A	11/1996	Takebayashi	5,982,906 A	11/1999	Ono
5,590,039 A	12/1996	Ikeda	5,983,190 A	11/1999	Trowerll
5,608,635 A	3/1997	Tamai	5,987,404 A	11/1999	DellaPietra
5,615,296 A	3/1997	Stanford	5,991,721 A	11/1999	Asano
5,617,407 A	4/1997	Bareis	5,995,119 A	11/1999	Cosatto
5,633,922 A	5/1997	August	5,995,928 A	11/1999	Nguyen
5,634,086 A	5/1997	Rtischev	5,995,943 A	11/1999	Bull
5,652,570 A	7/1997	Lepkofker	6,009,382 A	12/1999	Martino
5,675,629 A	10/1997	Raffel	6,014,559 A	1/2000	Amin
5,696,965 A	12/1997	Dedrisk	6,018,708 A	1/2000	Dahan
5,708,422 A	1/1998	Blonder	6,021,384 A	2/2000	Gorin
5,721,938 A	2/1998	Stuckey	6,028,514 A	2/2000	Lemelson
5,722,084 A	2/1998	Chakrin	6,035,267 A	3/2000	Watanabe
5,729,659 A	3/1998	Potter	6,044,347 A	3/2000	Abella
5,740,256 A	4/1998	CastelloDaCosta	6,049,602 A	4/2000	Foladare
5,742,763 A	4/1998	Jones	6,049,607 A	4/2000	Marash
5,748,841 A	5/1998	Morin	6,058,187 A	5/2000	Chen
5,748,974 A	5/1998	Johnson	6,067,513 A	5/2000	Ishimitsu
5,752,052 A	5/1998	Richardson	6,073,098 A	6/2000	Buchsbaum
5,754,784 A	5/1998	Garland	6,076,057 A	6/2000	Narayanan
5,761,631 A	6/1998	Nasukawa	6,076,059 A	6/2000	Glickman
5,774,841 A	6/1998	Salazar	6,078,886 A	6/2000	Dragosh
5,774,859 A	6/1998	Houser	6,081,774 A	6/2000	deHita
5,794,050 A	8/1998	Dahlgren	6,081,779 A	6/2000	Besling
5,794,196 A	8/1998	Yegnanarayanan	6,085,186 A	7/2000	Christianson
5,797,112 A	8/1998	Komatsu	6,101,241 A	8/2000	Boyce
5,799,276 A	8/1998	Komissarchik	6,108,631 A	8/2000	Ruhl
5,802,510 A	9/1998	Jones	6,119,087 A	9/2000	Kuhn
5,829,000 A	10/1998	Huang	6,119,101 A	9/2000	Peckover
			6,122,613 A	9/2000	Baker
			6,134,235 A	10/2000	Goldman
			6,144,667 A	11/2000	Doshi
			6,144,938 A	11/2000	Surace
			6,154,526 A	11/2000	Dahlke
			6,160,883 A	12/2000	Jackson
			6,167,377 A	12/2000	Gillick
			6,173,250 B1	1/2001	Jong
			6,173,266 B1	1/2001	Marx
			6,173,279 B1	1/2001	Levin
			6,175,858 B1	1/2001	Bulfer
			6,185,535 B1	2/2001	Hedin
			6,188,982 B1	2/2001	Chiang
			6,192,110 B1	2/2001	Abella
			6,192,338 B1	2/2001	Haszto
			6,195,634 B1	2/2001	Dudemaine
			6,195,651 B1	2/2001	Handel
			6,199,043 B1	3/2001	Happ
			6,208,964 B1	3/2001	Sabourin
			6,208,972 B1	3/2001	Grant
			6,219,346 B1	4/2001	Maxemchuk
			6,219,643 B1	4/2001	Cohen

US 11,087,385 B2

Page 3

(56)

References Cited

U.S. PATENT DOCUMENTS

6,219,645 B1	4/2001	Byers	6,539,348 B1	3/2003	Bond
6,226,612 B1	5/2001	Srenger	6,549,629 B2	4/2003	Finn
6,233,556 B1	5/2001	Teunen	6,553,372 B1	4/2003	Brassell
6,233,559 B1	5/2001	Balakrishnan	6,556,970 B1	4/2003	Sasaki
6,233,561 B1	5/2001	Junqua	6,556,973 B1	4/2003	Lewin
6,236,968 B1	5/2001	Kanevsky	6,560,576 B1	5/2003	Cohen
6,243,679 B1	6/2001	Mohri	6,560,590 B1	5/2003	Shwe
6,246,981 B1	6/2001	Papineni	6,567,778 B1	5/2003	ChaoChang
6,246,990 B1	6/2001	Happ	6,567,797 B1	5/2003	Schuetze
6,266,636 B1	7/2001	Kosaka	6,567,805 B1	5/2003	Johnson
6,269,336 B1	7/2001	Ladd	6,570,555 B1	5/2003	Prevost
6,272,455 B1	8/2001	Hoshen	6,570,964 B1	5/2003	Murweit
6,272,461 B1	8/2001	Meredith	6,571,279 B1	5/2003	Herz
6,275,231 B1	8/2001	Obradovich	6,574,597 B1	6/2003	Mohri
6,278,377 B1	8/2001	DeLine	6,574,624 B1	6/2003	Johnson
6,278,968 B1	8/2001	Franz	6,578,022 B1	6/2003	Foulger
6,286,002 B1	9/2001	Axaopoulos	6,581,103 B1	6/2003	Dengler
6,288,319 B1	9/2001	Catona	6,584,439 B1	6/2003	Geilhufe
6,292,767 B1	9/2001	Jackson	6,587,858 B1	7/2003	Strazza
6,301,560 B1	10/2001	Masters	6,591,185 B1	7/2003	Polidi
6,308,151 B1	10/2001	Smith	6,591,239 B1	7/2003	McCall
6,311,159 B1	10/2001	VanTichelen	6,594,257 B1	7/2003	Doshi
6,314,402 B1	11/2001	Monaco	6,594,367 B1	7/2003	Marash
6,321,196 B1	11/2001	Franceschi	6,598,018 B1	7/2003	Junqua
6,356,869 B1	3/2002	Chapados	6,601,026 B2	7/2003	Appelt
6,362,748 B1	3/2002	Huang	6,601,029 B1	7/2003	Pickering
6,366,882 B1	4/2002	Bijl	6,604,075 B1	8/2003	Brown
6,366,886 B1	4/2002	Dragosh	6,604,077 B2	8/2003	Dragosh
6,374,214 B1	4/2002	Friedland	6,606,598 B1	8/2003	Holthouse
6,374,226 B1	4/2002	Hunt	6,611,692 B2	8/2003	Raffel
6,377,913 B1	4/2002	Coffman	6,614,773 B1	9/2003	Maxemchuk
6,377,919 B1	4/2002	Burnett	6,615,172 B1	9/2003	Bennett
6,381,535 B1	4/2002	Durocher	6,622,119 B1	9/2003	Ramaswamy
6,385,596 B1	5/2002	Wiser	6,629,066 B1	9/2003	Jackson
6,385,646 B1	5/2002	Brown	6,631,346 B1	10/2003	Karaorman
6,389,398 B1	5/2002	Lustgarten	6,631,351 B1	10/2003	Ramachandran
6,393,403 B1	5/2002	Majaniemi	6,633,846 B1	10/2003	Bennett
6,393,428 B1	5/2002	Miller	6,636,790 B1	10/2003	Lightner
6,397,181 B1	5/2002	Li	6,643,620 B1	11/2003	Contolini
6,404,878 B1	6/2002	Jackson	6,647,363 B2	11/2003	Claassen
6,405,170 B1	6/2002	Phillips	6,650,747 B1	11/2003	Bala
6,408,272 B1	6/2002	White	6,658,388 B1	12/2003	Kleindienst
6,411,810 B1	6/2002	Maxemchuk	6,678,680 B1	1/2004	Woo
6,411,893 B2	6/2002	Ruhl	6,681,206 B1	1/2004	Gorin
6,415,257 B1	7/2002	Junqua	6,691,151 B1	2/2004	Cheyser
6,418,210 B1	7/2002	Sayko	6,701,294 B1	3/2004	Ball
6,420,975 B1	7/2002	DeLine	6,704,396 B2	3/2004	Parolkar
6,429,813 B2	8/2002	Feigen	6,704,576 B1	3/2004	Brachman
6,430,285 B1	8/2002	Bauer	6,704,708 B1	3/2004	Pickering
6,430,531 B1	8/2002	Polish	6,707,421 B1	3/2004	Drury
6,434,523 B1	8/2002	Monaco	6,708,150 B1	3/2004	Hirayama
6,434,524 B1	8/2002	Weber	6,721,001 B1	4/2004	Berstis
6,434,529 B1	8/2002	Walker	6,721,633 B2	4/2004	Funk
6,442,522 B1	8/2002	Carberry	6,721,706 B1	4/2004	Strubbe
6,446,114 B1	9/2002	Buffer	6,726,636 B2	4/2004	DerGhazarian
6,453,153 B1	9/2002	Bowker	6,732,088 B1	5/2004	Glance
6,453,292 B2	9/2002	Ramaswamy	6,735,592 B1	5/2004	Neumann
6,456,711 B1	9/2002	Cheung	6,739,556 B1	5/2004	Langston
6,456,974 B1	9/2002	Baker	6,741,931 B1	5/2004	Kohut
6,466,654 B1	10/2002	Cooper	6,742,021 B1	5/2004	Halverson
6,466,899 B1	10/2002	Yano	6,745,161 B1	6/2004	Arnold
6,470,315 B1	10/2002	Netsch	6,751,591 B1	6/2004	Gorin
6,487,494 B2	11/2002	Odinak	6,751,612 B1	6/2004	Schuetze
6,487,495 B1	11/2002	Gale	6,754,485 B1	6/2004	Obradovich
6,498,797 B1	12/2002	Anerousis	6,754,627 B2	6/2004	Woodward
6,499,013 B1	12/2002	Weber	6,754,647 B1	6/2004	Tackett
6,501,833 B2	12/2002	Phillips	6,757,362 B1	6/2004	Cooper
6,501,834 B1	12/2002	Milewski	6,757,544 B2	6/2004	Rangarajan
6,505,155 B1	1/2003	Vanbuskirk	6,757,718 B1	6/2004	Halverson
6,510,417 B1	1/2003	Woods	6,785,651 B1	8/2004	Wang
6,513,006 B2	1/2003	Howard	6,795,808 B1	9/2004	Strubbe
6,522,746 B1	2/2003	Marchok	6,801,604 B2	10/2004	Maes
6,523,061 B1	2/2003	Halverson	6,801,893 B1	10/2004	Backfried
6,526,335 B1	2/2003	Treyz	6,804,330 B1	10/2004	Jones
6,532,444 B1	3/2003	Weber	6,810,375 B1	10/2004	Ejerhed
			6,813,341 B1	11/2004	Mahoney
			6,816,830 B1	11/2004	Kempe
			6,823,308 B2	11/2004	Keiller
			6,829,603 B1	12/2004	Chai

US 11,087,385 B2

Page 4

(56)

References Cited

U.S. PATENT DOCUMENTS

6,832,230	B1	12/2004	Zilliacus	7,171,291	B2	1/2007	Obradovich
6,833,848	B1	12/2004	Wolff	7,174,300	B2	2/2007	Bush
6,839,896	B2	1/2005	Coffman	7,177,798	B2	2/2007	Hsu
6,850,603	B1	2/2005	Eberle	7,184,957	B2	2/2007	Brookes
6,856,990	B2	2/2005	Barile	7,190,770	B2	3/2007	Ando
6,859,776	B1 *	2/2005	Cohen	7,197,069	B2	3/2007	Agazzi
			H04M 3/493	7,197,460	B1	3/2007	Gupta
			704/270	7,203,644	B2	4/2007	Anderson
6,865,481	B2	3/2005	Kawazoe	7,206,418	B2	4/2007	Yang
6,868,380	B2	3/2005	Kroeker	7,207,011	B2	4/2007	Mulvey
6,868,385	B1	3/2005	Gerson	7,215,941	B2	5/2007	Beckmann
6,871,179	B1	3/2005	Kist	7,228,276	B2	6/2007	Omote
6,873,837	B1	3/2005	Yoshioka	7,231,343	B1	6/2007	Treadgold
6,877,001	B2	4/2005	Wolf	7,236,923	B1	6/2007	Gupta
6,877,134	B1	4/2005	Fuller	7,254,482	B2	8/2007	Kawasaki
6,882,970	B1	4/2005	Garner	7,272,212	B2	9/2007	Eberle
6,901,366	B1	5/2005	Kuhn	7,277,854	B2	10/2007	Bennett
6,910,003	B1	6/2005	Arnold	7,283,829	B2	10/2007	Christenson
6,912,498	B2	6/2005	Stevens	7,283,951	B2	10/2007	Marchisio
6,915,126	B2	7/2005	Mazzara, Jr.	7,289,606	B2	10/2007	Sibal
6,928,614	B1	8/2005	Everhart	7,299,186	B2	11/2007	Kuzunuki
6,934,756	B2	8/2005	Maes	7,301,093	B2	11/2007	Sater
6,937,977	B2	8/2005	Gerson	7,305,381	B1	12/2007	Poppink
6,937,982	B2	8/2005	Kitaoka	7,321,850	B2	1/2008	Wakita
6,941,266	B1	9/2005	Gorin	7,328,155	B2	2/2008	Endo
6,944,594	B2	9/2005	Busayapongchai	7,337,116	B2	2/2008	Charlesworth
6,950,821	B2	9/2005	Faybishenko	7,340,040	B1	3/2008	Saylor
6,954,755	B2	10/2005	Reisman	7,366,285	B2	4/2008	Parolkar
6,959,276	B2	10/2005	Droppo	7,366,669	B2	4/2008	Nishitani
6,961,700	B2	11/2005	Mitchell	7,376,586	B1	5/2008	Partovi
6,963,759	B1	11/2005	Gerson	7,376,645	B2	5/2008	Bernard
6,964,023	B2	11/2005	Maes	7,380,250	B2	5/2008	Schechter
6,968,311	B2	11/2005	Knockeart	7,386,443	B1	6/2008	Parthasarathy
6,973,387	B2	12/2005	Masclet	7,398,209	B2	7/2008	Kennewick
6,975,983	B1	12/2005	Fortescue	7,406,421	B2	7/2008	Odinak
6,975,993	B1	12/2005	Keiller	7,415,100	B2	8/2008	Cooper
6,980,092	B2	12/2005	Turnbull	7,415,414	B2	8/2008	Azara
6,983,055	B2	1/2006	Luo	7,421,393	B1	9/2008	DiFabrizio
6,990,513	B2	1/2006	Belfiore	7,424,431	B2	9/2008	Greene
6,996,531	B2	2/2006	Korall	7,447,635	B1	11/2008	Konopka
7,003,463	B1	2/2006	Maes	7,451,088	B1	11/2008	Ehlen
7,016,849	B2	3/2006	Arnold	7,454,368	B2	11/2008	Stillman
7,020,609	B2	3/2006	Thrift	7,454,608	B2	11/2008	Gopalakrishnan
7,024,364	B2	4/2006	Guerra	7,461,059	B2	12/2008	Richardson
7,027,586	B2	4/2006	Bushey	7,472,020	B2	12/2008	Brulle-Drews
7,027,974	B1	4/2006	Busch	7,472,060	B1	12/2008	Gorin
7,027,975	B1	4/2006	Pazandak	7,472,075	B2	12/2008	Odinak
7,031,908	B1	4/2006	Huang	7,477,909	B2	1/2009	Roth
7,035,415	B2	4/2006	Belt	7,478,036	B2	1/2009	Shen
7,036,128	B1	4/2006	Julia	7,487,088	B1	2/2009	Gorin
7,043,425	B2	5/2006	Pao	7,487,110	B2	2/2009	Bennett
7,054,817	B2	5/2006	Shao	7,493,259	B2	2/2009	Jones
7,058,890	B2	6/2006	George	7,493,559	B1	2/2009	Wolff
7,062,488	B1	6/2006	Reisman	7,502,672	B1	3/2009	Kolls
7,069,220	B2	6/2006	Coffman	7,502,730	B2	3/2009	Wang
7,072,834	B2	7/2006	Zhou	7,502,738	B2	3/2009	Kennewick
7,072,888	B1	7/2006	Perkins	7,512,906	B1	3/2009	Baier
7,076,362	B2	7/2006	Ohtsuji	7,516,076	B2	4/2009	Walker
7,082,469	B2	7/2006	Gold	7,529,675	B2	5/2009	Maes
7,085,708	B2	8/2006	Manson	7,536,297	B2	5/2009	Byrd
7,092,928	B1	8/2006	Elad	7,536,374	B2	5/2009	Au
7,107,210	B2	9/2006	Deng	7,542,894	B2	6/2009	Murata
7,107,218	B1	9/2006	Preston	7,546,382	B2	6/2009	Healey
7,110,951	B1	9/2006	Lemelson	7,548,491	B2	6/2009	Macfarlane
7,127,395	B1	10/2006	Gorin	7,552,054	B1	6/2009	Stifelman
7,127,400	B2	10/2006	Koch	7,558,730	B2	7/2009	Davis
7,130,390	B2	10/2006	Abhuri	7,574,362	B2	8/2009	Walker
7,136,875	B2	11/2006	Anderson	7,577,244	B2	8/2009	Taschereau
7,137,126	B1	11/2006	Coffman	7,606,708	B2	10/2009	Hwang
7,143,037	B1	11/2006	Chestnut	7,606,712	B1	10/2009	Smith
7,143,039	B1	11/2006	Stifelman	7,620,549	B2	11/2009	DiCristo
7,146,315	B2	12/2006	Balan	7,634,409	B2	12/2009	Kennewick
7,146,319	B2	12/2006	Hunt	7,640,006	B2	12/2009	Portman
7,149,696	B2	12/2006	Shimizu	7,640,160	B2	12/2009	DiCristo
7,165,028	B2	1/2007	Gong	7,640,272	B2	12/2009	Mahajan
7,170,993	B2	1/2007	Anderson	7,672,931	B2	3/2010	Hurst-Hiller
				7,676,365	B2	3/2010	Hwang
				7,676,369	B2	3/2010	Fujimoto
				7,684,977	B2	3/2010	Morikawa

US 11,087,385 B2

Page 5

(56)

References Cited

U.S. PATENT DOCUMENTS

7,693,720	B2	4/2010	Kennewick	8,509,403	B2	8/2013	Chiu
7,697,673	B2	4/2010	Chiu	8,515,765	B2	8/2013	Baldwin
7,706,616	B2	4/2010	Kristensson	8,527,274	B2	9/2013	Freeman
7,729,913	B1	6/2010	Lee	8,577,671	B1	11/2013	Barve
7,729,916	B2	6/2010	Coffman	8,589,161	B2	11/2013	Kennewick
7,729,918	B2	6/2010	Walker	8,612,205	B2	12/2013	Hanneman
7,729,920	B2	6/2010	Chaar	8,612,206	B2	12/2013	Chalabi
7,734,287	B2	6/2010	Ying	8,620,659	B2	12/2013	DiCristo
7,748,021	B2	6/2010	Obradovich	8,719,005	B1	5/2014	Lee
7,774,333	B2	8/2010	Colledge	8,719,009	B2	5/2014	Baldwin
7,788,084	B2	8/2010	Brun	8,719,026	B2	5/2014	Kennewick
7,792,257	B1	9/2010	Vanier	8,731,929	B2	5/2014	Kennewick
7,801,731	B2	9/2010	Odinak	8,738,380	B2	5/2014	Baldwin
7,809,570	B2	10/2010	Kennewick	8,849,652	B2	9/2014	Weider
7,818,176	B2	10/2010	Freeman	8,849,670	B2	9/2014	DiCristo
7,831,426	B2	11/2010	Bennett	8,849,696	B2	9/2014	Pansari
7,831,433	B1	11/2010	Belvin	8,849,791	B1	9/2014	Hertschuh
7,856,358	B2	12/2010	Ho	8,886,536	B2	11/2014	Freeman
7,873,519	B2	1/2011	Bennett	8,972,243	B1	3/2015	Strom
7,873,523	B2	1/2011	Potter	8,972,354	B1	3/2015	Telang
7,873,654	B2	1/2011	Bernard	8,983,839	B2	3/2015	Kennewick
7,881,936	B2	2/2011	Longe	9,009,046	B1	4/2015	Stewart
7,890,324	B2	2/2011	Bangalore	9,015,049	B2	4/2015	Baldwin
7,894,849	B2	2/2011	Kass	9,037,455	B1	5/2015	Faaborg
2,902,969	A1	3/2011	Obradovich	9,070,366	B1	6/2015	Mathias
7,917,367	B2	3/2011	DiCristo	9,070,367	B1	6/2015	Hoffmeister
7,920,682	B2	4/2011	Byrne	9,105,266	B2	8/2015	Baldwin
7,949,529	B2	5/2011	Weider	9,171,541	B2	10/2015	Kennewick
7,949,537	B2	5/2011	Walker	9,269,097	B2	2/2016	Freeman
7,953,732	B2	5/2011	Frank	9,305,548	B2	4/2016	Kennewick
7,974,875	B1	7/2011	Quilici	9,308,445	B1	4/2016	Merzenich
7,983,917	B2	7/2011	Kennewick	9,318,108	B2	4/2016	Gruber
7,984,287	B2	7/2011	Gopalakrishnan	9,406,078	B2	8/2016	Freeman
8,005,683	B2	8/2011	Tessel	9,443,514	B1	9/2016	Taubman
8,015,006	B2	9/2011	Kennewick	9,502,025	B2	11/2016	Kennewick
8,024,186	B1	9/2011	De Bonet	9,626,703	B2	4/2017	Kennewick, Sr.
8,027,965	B2	9/2011	Takehara	9,711,143	B2	7/2017	Kennewick
8,032,383	B1	10/2011	Bhardwaj	10,089,984	B2	10/2018	Kennewick
8,060,367	B2	11/2011	Keaveney	2001/0011336	A1	8/2001	Sitka
8,069,046	B2	11/2011	Kennewick	2001/0014857	A1	8/2001	Wang
8,073,681	B2	12/2011	Baldwin	2001/0021905	A1	9/2001	Burnett
8,077,975	B2	12/2011	Ma	2001/0039492	A1	11/2001	Nemoto
8,078,502	B2 *	12/2011	Hao G06Q 30/0641 705/26.61	2001/0041980	A1	11/2001	Howard
8,082,153	B2	12/2011	Coffman	2001/0047261	A1	11/2001	Kassan
8,086,463	B2	12/2011	Ativanichayaphong	2001/0049601	A1	12/2001	Kroeker
8,103,510	B2	1/2012	Sato	2001/0054087	A1	12/2001	Flom
8,112,275	B2	2/2012	Kennewick	2002/0002548	A1	1/2002	Roundtree
8,140,327	B2	3/2012	Kennewick	2002/0007267	A1	1/2002	Batchilo
8,140,335	B2	3/2012	Kennewick	2002/0010584	A1	1/2002	Schultz
8,145,489	B2	3/2012	Freeman	2002/0015500	A1	2/2002	Belt
8,150,694	B2	4/2012	Kennewick	2002/0022927	A1	2/2002	Lemelson
8,155,962	B2	4/2012	Kennewick	2002/0022956	A1	2/2002	Ukrainczyk
8,170,867	B2	5/2012	Germain	2002/0029186	A1	3/2002	Roth
8,180,037	B1	5/2012	Delker	2002/0029261	A1	3/2002	Shibata
8,190,436	B2	5/2012	Bangalore	2002/0032752	A1	3/2002	Gold
8,195,468	B2	6/2012	Weider	2002/0035501	A1	3/2002	Handel
8,200,485	B1	6/2012	Lee	2002/0040297	A1	4/2002	Tsiao
8,204,751	B1	6/2012	Di Fabbri	2002/0049535	A1	4/2002	Rigo
8,219,399	B2	7/2012	Lutz	2002/0049805	A1	4/2002	Yamada
8,219,599	B2	7/2012	Tunstall-Pedoe	2002/0059068	A1	5/2002	Rose
8,224,652	B2	7/2012	Wang	2002/0065568	A1	5/2002	Silfvast
8,255,224	B2	8/2012	Singleton	2002/0067839	A1	6/2002	Heinrich
8,326,599	B2	12/2012	Tomeh	2002/0069059	A1	6/2002	Smith
8,326,627	B2	12/2012	Kennewick	2002/0069071	A1	6/2002	Knockeart
8,326,634	B2	12/2012	DiCristo	2002/0073176	A1	6/2002	Ikeda
8,326,637	B2	12/2012	Baldwin	2002/0082911	A1	6/2002	Dunn
8,332,224	B2	12/2012	DiCristo	2002/0087312	A1	7/2002	Lee
8,340,975	B1	12/2012	Rosenberger	2002/0087326	A1	7/2002	Lee
8,346,563	B1	1/2013	Hjelm	2002/0087525	A1	7/2002	Abbott
8,370,147	B2	2/2013	Kennewick	2002/0107694	A1	8/2002	Lerg
8,447,607	B2	5/2013	Weider	2002/0120609	A1	8/2002	Lang
8,447,651	B1	5/2013	Scholl	2002/0124050	A1	9/2002	Middeljans
8,452,598	B2	5/2013	Kennewick	2002/0133347	A1	9/2002	Schoneburg
8,503,995	B2	8/2013	Ramer	2002/0133354	A1	9/2002	Ross
				2002/0133402	A1	9/2002	Faber
				2002/0135618	A1	9/2002	Maes
				2002/0138248	A1	9/2002	Corston-Oliver
				2002/0143532	A1	10/2002	McLean
				2002/0143535	A1	10/2002	Kist

US 11,087,385 B2

Page 6

(56)

References Cited

U.S. PATENT DOCUMENTS

2002/0152260	A1	10/2002	Chen	2004/0199375	A1	10/2004	Ehsani
2002/0161646	A1	10/2002	Gailey	2004/0199387	A1	10/2004	Wang
2002/0161647	A1	10/2002	Galley	2004/0199389	A1	10/2004	Geiger
2002/0169597	A1	11/2002	Fain	2004/0201607	A1	10/2004	Mulvey
2002/0173333	A1	11/2002	Buchholz	2004/0204939	A1	10/2004	Liu
2002/0173961	A1	11/2002	Guerra	2004/0205671	A1	10/2004	Sukehiro
2002/0178003	A1	11/2002	Gehrke	2004/0220800	A1	11/2004	Kong
2002/0184373	A1	12/2002	Maes	2004/0243393	A1	12/2004	Wang
2002/0188455	A1	12/2002	Shioda	2004/0243417	A9	12/2004	Pitts
2002/0188602	A1	12/2002	Stubler	2004/0247092	A1	12/2004	Timmins
2002/0198714	A1	12/2002	Zhou	2004/0249636	A1	12/2004	Applebaum
2003/0005033	A1	1/2003	Mohan	2005/0015256	A1	1/2005	Kargman
2003/0014261	A1	1/2003	Kageyama	2005/0021331	A1	1/2005	Huang
2003/0016835	A1	1/2003	Elko	2005/0021334	A1	1/2005	Iwahashi
2003/0036903	A1	2/2003	Konopka	2005/0021470	A1	1/2005	Martin
2003/0046071	A1	3/2003	Wyman	2005/0021826	A1	1/2005	Kumar
2003/0046281	A1	3/2003	Son	2005/0033574	A1	2/2005	Kim
2003/0046346	A1	3/2003	Mumick	2005/0033582	A1	2/2005	Gadd
2003/0064709	A1	4/2003	Galley	2005/0043940	A1	2/2005	Elder
2003/0065427	A1	4/2003	Funk	2005/0043956	A1	2/2005	Aoyama
2003/0069734	A1	4/2003	Everhart	2005/0060142	A1	3/2005	Visser
2003/0069880	A1	4/2003	Harrison	2005/0080632	A1	4/2005	Endo
2003/0088421	A1	5/2003	Maes	2005/0080821	A1	4/2005	Breil
2003/0093419	A1	5/2003	Bangalore	2005/0102282	A1	5/2005	Linden
2003/0097249	A1	5/2003	Walker	2005/0114116	A1	5/2005	Fiedler
2003/0110037	A1	6/2003	Walker	2005/0125232	A1	6/2005	Gadd
2003/0112267	A1	6/2003	Belrose	2005/0131673	A1	6/2005	Koizumi
2003/0115062	A1	6/2003	Walker	2005/0137850	A1	6/2005	Odell
2003/0120493	A1	6/2003	Gupta	2005/0137877	A1	6/2005	Oesterling
2003/0135488	A1	7/2003	Amir	2005/0138168	A1	6/2005	Hoffman
2003/0144846	A1	7/2003	Denenberg	2005/0143994	A1	6/2005	Mori
2003/0154082	A1	8/2003	Toguri	2005/0144013	A1	6/2005	Fujimoto
2003/0158731	A1	8/2003	Falcon	2005/0144187	A1	6/2005	Che
2003/0161448	A1	8/2003	Parolkar	2005/0149319	A1	7/2005	Honda
2003/0167167	A1	9/2003	Gong	2005/0169441	A1	8/2005	Yacoub
2003/0174155	A1	9/2003	Weng	2005/0216254	A1	9/2005	Gupta
2003/0182123	A1	9/2003	Mitsuyoshi	2005/0222763	A1	10/2005	Uyeki
2003/0182132	A1	9/2003	Niemoeller	2005/0234637	A1	10/2005	Obradovich
2003/0187643	A1	10/2003	VanThong	2005/0234727	A1	10/2005	Chiu
2003/0204492	A1	10/2003	Wolf	2005/0240412	A1	10/2005	Fujita
2003/0206640	A1	11/2003	Malvar	2005/0246174	A1	11/2005	DeGolia
2003/0212550	A1	11/2003	Ubale	2005/0283364	A1	12/2005	Longe
2003/0212558	A1	11/2003	Matula	2005/0283532	A1	12/2005	Kim
2003/0212562	A1	11/2003	Patel	2005/0283752	A1	12/2005	Fruchter
2003/0216919	A1	11/2003	Roushar	2006/0041431	A1	2/2006	Maes
2003/0225825	A1	12/2003	Healey	2006/0046740	A1	3/2006	Johnson
2003/0233230	A1	12/2003	Ammicht	2006/0047509	A1	3/2006	Ding
2003/0236664	A1	12/2003	Sharma	2006/0072738	A1	4/2006	Louis
2004/0006475	A1	1/2004	Ehlen	2006/0074670	A1	4/2006	Weng
2004/0010358	A1	1/2004	Oesterling	2006/0074671	A1	4/2006	Farmaner
2004/0025115	A1	2/2004	Sienel	2006/0080098	A1	4/2006	Campbell
2004/0030741	A1	2/2004	Wolton	2006/0100851	A1	5/2006	Schonebeck
2004/0036601	A1	2/2004	Obradovich	2006/0106769	A1	5/2006	Gibbs
2004/0044516	A1	3/2004	Kennewick	2006/0129409	A1	6/2006	Mizutani
2004/0093567	A1	5/2004	Schabes	2006/0130002	A1	6/2006	Hirayama
2004/0098245	A1	5/2004	Walker	2006/0149633	A1	7/2006	Voisin
2004/0102977	A1	5/2004	Metzler	2006/0182085	A1	8/2006	Sweeney
2004/0117179	A1	6/2004	Balasuriya	2006/0206310	A1	9/2006	Ravikumar
2004/0117804	A1	6/2004	Scahill	2006/0217133	A1	9/2006	Christenson
2004/0122673	A1	6/2004	Park	2006/0236343	A1	10/2006	Chang
2004/0122674	A1	6/2004	Bangalore	2006/0242017	A1	10/2006	Libes
2004/0128514	A1	7/2004	Rhoads	2006/0247919	A1	11/2006	Specht
2004/0133793	A1	7/2004	Ginter	2006/0253247	A1	11/2006	de Silva
2004/0140989	A1	7/2004	Papageorge	2006/0253281	A1	11/2006	Letzt
2004/0143440	A1	7/2004	Prasad	2006/0285662	A1	12/2006	Yin
2004/0148154	A1	7/2004	Acero	2007/0011159	A1	1/2007	Hillis
2004/0148170	A1	7/2004	Acero	2007/0033005	A1	2/2007	Cristo
2004/0158555	A1	8/2004	Seedman	2007/0033020	A1	2/2007	Francois
2004/0166832	A1	8/2004	Portman	2007/0033526	A1	2/2007	Thompson
2004/0167771	A1	8/2004	Duan	2007/0038436	A1	2/2007	Cristo
2004/0172247	A1	9/2004	Yoon	2007/0038445	A1	2/2007	Helbing
2004/0172258	A1	9/2004	Dominach	2007/0043569	A1	2/2007	Potter
2004/0189697	A1	9/2004	Fukuoka	2007/0043574	A1	2/2007	Coffman
2004/0193408	A1	9/2004	Hunt	2007/0043868	A1	2/2007	Kumar
2004/0193420	A1	9/2004	Kennewick	2007/0050191	A1	3/2007	Weider
				2007/0050279	A1	3/2007	Huang
				2007/0055525	A1	3/2007	Kennewick
				2007/0060114	A1	3/2007	Ramer
				2007/0061067	A1	3/2007	Zeinstra

US 11,087,385 B2

Page 7

(56) References Cited

U.S. PATENT DOCUMENTS

2007/0061735	A1	3/2007	Hoffberg	2009/0030686	A1	1/2009	Weng	
2007/0067310	A1	3/2007	Gupta	2009/0052635	A1	2/2009	Jones	
2007/0073544	A1	3/2007	Millett	2009/0055176	A1	2/2009	Hu	
2007/0078708	A1	4/2007	Yu	2009/0067599	A1	3/2009	Agarwal	
2007/0078709	A1	4/2007	Rajaram	2009/0076827	A1	3/2009	Bulitta	
2007/0078814	A1	4/2007	Flowers	2009/0106029	A1	4/2009	DeLine	
2007/0094003	A1	4/2007	Huang	2009/0117885	A1	5/2009	Roth	
2007/0100797	A1	5/2007	Thun	2009/0144131	A1	6/2009	Chiu	
2007/0106499	A1	5/2007	Dahlgren	2009/0144271	A1	6/2009	Richardson	
2007/0112555	A1	5/2007	Lavi	2009/0150156	A1	6/2009	Kennewick	
2007/0112630	A1	5/2007	Lau	2009/0157382	A1	6/2009	Bar	
2007/0118357	A1	5/2007	Kasravi	2009/0164216	A1	6/2009	Chengalvarayan	
2007/0124057	A1	5/2007	Prieto	2009/0171664	A1	7/2009	Kennewick	
2007/0135101	A1	6/2007	Ramati	2009/0171912	A1	7/2009	Nash	
2007/0146833	A1	6/2007	Satomi	2009/0197582	A1	8/2009	Lewis	
2007/0162296	A1	7/2007	Altberg	2009/0216540	A1	8/2009	Tessel	
2007/0174258	A1	7/2007	Jones	2009/0248565	A1	10/2009	Chuang	
2007/0179778	A1	8/2007	Gong	2009/0248605	A1	10/2009	Mitchell	
2007/0185859	A1	8/2007	Flowers	2009/0259561	A1 *	10/2009	Boys	G06Q 30/06 705/26.1
2007/0186165	A1	8/2007	Maislos	2009/0259646	A1	10/2009	Fujita	
2007/0192309	A1	8/2007	Fischer	2009/0265163	A1	10/2009	Li	
2007/0198267	A1	8/2007	Jones	2009/0271194	A1	10/2009	Davis	
2007/0203699	A1	8/2007	Nagashima	2009/0273563	A1	11/2009	Pryor	
2007/0203736	A1	8/2007	Ashton	2009/0276700	A1	11/2009	Anderson	
2007/0208732	A1	9/2007	Flowers	2009/0287680	A1	11/2009	Paek	
2007/0214182	A1	9/2007	Rosenberg	2009/0299745	A1	12/2009	Kennewick	
2007/0250901	A1	10/2007	McIntire	2009/0299857	A1	12/2009	Brubaker	
2007/0265850	A1	11/2007	Kennewick	2009/0304161	A1	12/2009	Pettyjohn	
2007/0266257	A1	11/2007	Camaisa	2009/0307031	A1	12/2009	Winkler	
2007/0276651	A1	11/2007	Bliss	2009/0313026	A1	12/2009	Coffman	
2007/0294615	A1	12/2007	Sathe	2009/0319517	A1	12/2009	Guha	
2007/0299824	A1	12/2007	Pan	2010/0023320	A1	1/2010	Cristo	
2008/0010135	A1	1/2008	Schrock	2010/0023331	A1	1/2010	Duta	
2008/0014908	A1	1/2008	Vasant	2010/0029261	A1	2/2010	Mikkelsen	
2008/0032622	A1	2/2008	Kopra	2010/0036967	A1	2/2010	Caine	
2008/0034032	A1	2/2008	Healey	2010/0049501	A1	2/2010	Kennewick	
2008/0046311	A1	2/2008	Shahine	2010/0049514	A1	2/2010	Kennewick	
2008/0059188	A1	3/2008	Konopka	2010/0057443	A1	3/2010	Cristo	
2008/0065386	A1	3/2008	Cross	2010/0063880	A1	3/2010	Atsmon	
2008/0065389	A1	3/2008	Cross	2010/0064025	A1	3/2010	Nelimarkka	
2008/0065390	A1	3/2008	Ativanichayaphong	2010/0076778	A1	3/2010	Kondrk	
2008/0086455	A1	4/2008	Meisels	2010/0094707	A1	4/2010	Freer	
2008/0091406	A1	4/2008	Baldwin	2010/0138300	A1	6/2010	Wallis	
2008/0103761	A1	5/2008	Printz	2010/0145700	A1	6/2010	Kennewick	
2008/0103781	A1	5/2008	Wasson	2010/0185512	A1	7/2010	Borger	
2008/0104071	A1	5/2008	Pragada	2010/0191856	A1	7/2010	Gupta	
2008/0109285	A1	5/2008	Reuther	2010/0204986	A1	8/2010	Kennewick	
2008/0115163	A1	5/2008	Gilboa	2010/0204994	A1	8/2010	Kennewick	
2008/0126091	A1	5/2008	Clark	2010/0217604	A1	8/2010	Baldwin	
2008/0126284	A1	5/2008	Forbes	2010/0268536	A1	10/2010	Suendermann	
2008/0133215	A1	6/2008	Sarukkai	2010/0286985	A1	11/2010	Kennewick	
2008/0140385	A1	6/2008	Mahajan	2010/0299142	A1	11/2010	Freeman	
2008/0147396	A1	6/2008	Wang	2010/0312547	A1	12/2010	Van Os	
2008/0147410	A1	6/2008	Odinak	2010/0312566	A1	12/2010	Odinak	
2008/0147637	A1	6/2008	Li	2010/0318357	A1	12/2010	Istvan	
2008/0154604	A1	6/2008	Sathish	2010/0331064	A1	12/2010	Michelstein	
2008/0162471	A1	7/2008	Bernard	2011/0022393	A1	1/2011	Waller	
2008/0177530	A1	7/2008	Cross	2011/0106527	A1	5/2011	Chiu	
2008/0184164	A1	7/2008	Di Fabbizio	2011/0112827	A1	5/2011	Kennewick	
2008/0189110	A1	8/2008	Freeman	2011/0112921	A1	5/2011	Kennewick	
2008/0189187	A1	8/2008	Hao	2011/0119049	A1	5/2011	Ylonen	
2008/0228496	A1	9/2008	Yu	2011/0131036	A1	6/2011	DiCristo	
2008/0235023	A1	9/2008	Kennewick	2011/0131045	A1	6/2011	Cristo	
2008/0235027	A1	9/2008	Cross	2011/0231182	A1	9/2011	Weider	
2008/0269958	A1	10/2008	Filev	2011/0231188	A1	9/2011	Kennewick	
2008/0270135	A1	10/2008	Goel	2011/0238409	A1	9/2011	Larcheveque	
2008/0270224	A1	10/2008	Portman	2011/0307167	A1	12/2011	Taschereau	
2008/0294437	A1	11/2008	Nakano	2012/0022857	A1	1/2012	Baldwin	
2008/0294994	A1	11/2008	Kruger	2012/0041753	A1	2/2012	Dymetman	
2008/0306743	A1	12/2008	Di Fabbizio	2012/0046935	A1	2/2012	Nagao	
2008/0319751	A1	12/2008	Kennewick	2012/0101809	A1	4/2012	Kennewick	
2009/0006077	A1	1/2009	Keaveney	2012/0101810	A1	4/2012	Kennewick	
2009/0006194	A1	1/2009	Sridharan	2012/0109753	A1	5/2012	Kennewick	
2009/0018829	A1	1/2009	Kuperstein	2012/0150620	A1	6/2012	Mandyam	
2009/0024476	A1	1/2009	Baar	2012/0150636	A1	6/2012	Freeman	
				2012/0239498	A1	9/2012	Ramer	
				2012/0240060	A1	9/2012	Pennington	
				2012/0265528	A1	10/2012	Gruber	

US 11,087,385 B2

Page 8

(56) References Cited

U.S. PATENT DOCUMENTS

2012/0278073	A1	11/2012	Weider	
2013/0006734	A1	1/2013	Ocko	
2013/0054228	A1	2/2013	Baldwin	
2013/0060625	A1	3/2013	Davis	
2013/0080177	A1	3/2013	Chen	
2013/0211710	A1	8/2013	Kennewick	
2013/0253929	A1	9/2013	Weider	
2013/0254314	A1	9/2013	Chow	
2013/0297293	A1	11/2013	Cristo	
2013/0304473	A1	11/2013	Baldwin	
2013/0311324	A1	11/2013	Stoll	
2013/0332454	A1	12/2013	Stuhec	
2013/0339022	A1	12/2013	Baldwin	
2014/0006951	A1	1/2014	Hunter	
2014/0012577	A1	1/2014	Freeman	
2014/0025371	A1	1/2014	Min	
2014/0108013	A1	4/2014	Cristo	
2014/0136187	A1	5/2014	Wolverton	
2014/0136259	A1*	5/2014	Kinsey, II	 G06Q 30/0631 705/7.16
2014/0156278	A1	6/2014	Kennewick	
2014/0195238	A1	7/2014	Terao	
2014/0236575	A1	8/2014	Tur	
2014/0249821	A1	9/2014	Kennewick	
2014/0249822	A1	9/2014	Baldwin	
2014/0278413	A1	9/2014	Pitschel	
2014/0278416	A1	9/2014	Schuster	
2014/0288934	A1	9/2014	Kennewick	
2014/0330552	A1	11/2014	Bangalore	
2014/0337007	A1	11/2014	Waibel	
2014/0365222	A1	12/2014	Weider	
2015/0019211	A1	1/2015	Simard	
2015/0019217	A1	1/2015	Cristo	
2015/0019227	A1	1/2015	Anandarajah	
2015/0066479	A1	3/2015	Pasupalak	
2015/0066627	A1	3/2015	Freeman	
2015/0073910	A1	3/2015	Kennewick	
2015/0095159	A1	4/2015	Kennewick	
2015/0142447	A1	5/2015	Kennewick	
2015/0170641	A1	6/2015	Kennewick	
2015/0193379	A1	7/2015	Mehta	
2015/0199339	A1	7/2015	Mirkin	
2015/0228276	A1	8/2015	Baldwin	
2015/0293917	A1	10/2015	Bufe	
2015/0348544	A1	12/2015	Baldwin	
2015/0348551	A1	12/2015	Gruber	
2015/0364133	A1	12/2015	Freeman	
2016/0049152	A1	2/2016	Kennewick	
2016/0078482	A1	3/2016	Kennewick	
2016/0078491	A1	3/2016	Kennewick	
2016/0078504	A1	3/2016	Kennewick	
2016/0078773	A1	3/2016	Carter	
2016/0110347	A1	4/2016	Kennewick	
2016/0148610	A1	5/2016	Kennewick	
2016/0148612	A1	5/2016	Guo	
2016/0188292	A1	6/2016	Carter	
2016/0188573	A1	6/2016	Tang	
2016/0217785	A1	7/2016	Kennewick	
2016/0335676	A1	11/2016	Freeman	
2017/0004588	A1	1/2017	Isaacson	

FOREIGN PATENT DOCUMENTS

CN	101236635	A	8/2008
CN	103229206	A	7/2013
EP	1320043	A2	6/2003
EP	1646037		4/2006
FR	2867583		9/2005
JP	H08263258		10/1996
JP	H11249773		9/1999
JP	2001071289		3/2001
JP	2002297626		10/2002
JP	2006146881		6/2006
JP	2008027454		2/2008

JP	2008058465	3/2008
JP	2008139928	6/2008
JP	2011504304	2/2011
JP	2012518847	8/2012
WO	9946763	9/1999
WO	0021232	4/2000
WO	0046792	8/2000
WO	0129742	4/2001
WO	0129742	A2 4/2001
WO	0171609	A2 9/2001
WO	0178065	10/2001
WO	2004072954	8/2004
WO	2005010702	A2 2/2005
WO	2007019318	1/2007
WO	2007021587	1/2007
WO	2007027546	1/2007
WO	2007027989	1/2007
WO	2008098039	1/2008
WO	2008118195	1/2008
WO	2009075912	1/2009
WO	2009145796	1/2009
WO	2009111721	9/2009
WO	2010096752	1/2010
WO	2016044290	3/2016
WO	2016044316	3/2016
WO	2016044319	3/2016
WO	2016044321	3/2016
WO	2016061309	4/2016

OTHER PUBLICATIONS

"Statement in Accordance with the Notice from the European Patent Office" dated Oct. 1, 2007 Concerning Business Methods (OJ EPO Nov. 2007, 592-593), XP002456252.

Arrington, Michael, "Google Redefines GPS Navigation Landscape: Google Maps Navigation for Android 2.0", TechCrunch, printed from the Internet <<http://www.techcrunch.com/2009/10/28/google-redefines-car-gps-navigation-google-maps-navigation-android/>>, Oct. 28, 2009, 4 pages.

Bazzi, Issam et al., "Heterogeneous Lexical Units for Automatic Speech Recognition: Preliminary Investigations", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, Jun. 5-9, 2000, XP010507574, pp. 1257-1260.

Belvin, Robert, et al., "Development of the HRL Route Navigation Dialogue System", Proceedings of the First International Conference on Human Language Technology Research, San Diego, 2001, pp. 1-5.

Chai et al., "MIND: A Semantics-Based Multimodal Interpretation Framework for Conversational Systems", Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Jun. 2002, pp. 37-46.

Cheyen et al., "Multimodal Maps: An Agent-Based Approach", International Conference on Cooperative Multimodal Communication (CMC/95), May 24-26, 1995, pp. 111-121.

Davis, Z., et al., A Personal Handheld Multi-Modal Shopping Assistant, IEEE, 2006, 9 pages.

El Meliani et al., "A Syllabic-Filler-Based Continuous Speech Recognizer for Unlimited Vocabulary", Canadian Conference on Electrical and Computer Engineering, vol. 2, Sep. 5-8, 1995, pp. 1007-1010.

Elio et al., "On Abstract Task Models and Conversation Policies" in Workshop on Specifying and Implementing Conversation Policies, Autonomous Agents '99, Seattle, 1999, 10 pages.

Kirchhoff, Katrin, "Syllable-Level Desynchronisation of Phonetic Features for Speech Recognition", Proceedings of the Fourth International Conference on Spoken Language, 1996, ICSLP 96, vol. 4, IEEE, 1996, 3 pages.

Kuhn, Thomas, et al., "Hybrid In-Car Speech Recognition for Mobile Multimedia Applications", Vehicular Technology Conference, IEEE, Jul. 1999, pp. 2009-2013.

Lin, Bor-shen, et al., "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History", ASRU'99, 1999, 4 pages.

US 11,087,385 B2

Page 9

(56)

References Cited

OTHER PUBLICATIONS

Lind, R., et al., "The Network Vehicle—A Glimpse into the Future of Mobile Multi-Media", IEEE Aerosp. Electron. Systems Magazine, vol. 14, No. 9, Sep. 1999, pp. 27-32.

Mao, Mark Z., "Automatic Training Set Segmentation for Multi-Pass Speech Recognition", Department of Electrical Engineering, Stanford University, CA, copyright 2005, IEEE, pp. I-685 to I-688.

O'Shaughnessy, Douglas, "Interacting with Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, vol. 91, No. 9, Sep. 1, 2003, XP011100665, pp. 1272-1305.

Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., Jul. 28, 2002, 1 page.

Turunen, "Adaptive Interaction Methods in Speech User Interfaces", Conference on Human Factors in Computing Systems; Seattle, Washington, 2001, pp. 91-92.

Vanhoecke, Vincent, "Confidence Scoring and Rejection Using Multi-Pass Speech Recognition", Nuance Communications, Menlo Park, CA, 2005, 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, 1998, 4 pages.

Wu, Su-Lin, et al., "Incorporating Information from Syllable-Length Time Scales into Automatic Speech Recognition", Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, vol. 2, IEEE, 1998, 4 pages.

Wu, Su-Lin, et al., "Integrating Syllable Boundary Information into Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, 1997, vol. 2, IEEE, 1997, 4 pages.

Zhao, Yilin, "Telematics: Safe and Fun Driving", IEEE Intelligent Systems, vol. 17, Issue 1, 2002, pp. 10-14.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01346, Decision Granting Institution of Inter Partes Review of U.S. Pat. No. 9,015,049 B2, dated Feb. 4, 2021, 24 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01367, Decision Granting Institution of Inter Partes Review of U.S. Pat. No. 8,073,681 B2, dated Mar. 11, 2021, 22 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01374, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 8,886,536 B2, dated Feb. 4, 2021, 37 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01377, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 8,886,536 B2, dated Feb. 4, 2021, 29 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01380, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 9,626,703 B2, dated Feb. 4, 2021, 28 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01381, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 9,626,703 B2, dated Feb. 4, 2021, 29 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01388, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 9,269,097 B2, dated Feb. 24, 2021, 25 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. *VB Assets, LLC*, IPR2020-01390, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 7,818,176, dated Mar. 11, 2021, 27 pages.

Asthana, A., et al., "A Small Domain Communications System for Personalized Shopping Assistance", Proceedings of 1994 International Conference on Personal Wireless Communications, IEEE Press, Aug. 1994, pp. 199-203.

Huang et al., "Spoken Language Processing: A Guide to Theory, Algorithm, and System Development", Prentice Hall, 2001, 1010 pages.

IEEE 100 The Authoritative Dictionary of IEEE Standards Terms, Seventh Edition, Standards Information Network, IEEE Press, Print ISBN 0-7381-2601-2, Published Dec. 2000, 3 pages.

Lucente, Mark, "Conversational Interfaces for E-Commerce Applications", Communications of the ACM, vol. 43, No. 9, Sep. 2000, pp. 59-61.

Senef, Stephanie, et al., "Hypothesis Selection and Resolution in the Mercury Flight Reservation System", Spoken Language Systems Group, MIT, 2001, 8 pages.

* cited by examiner

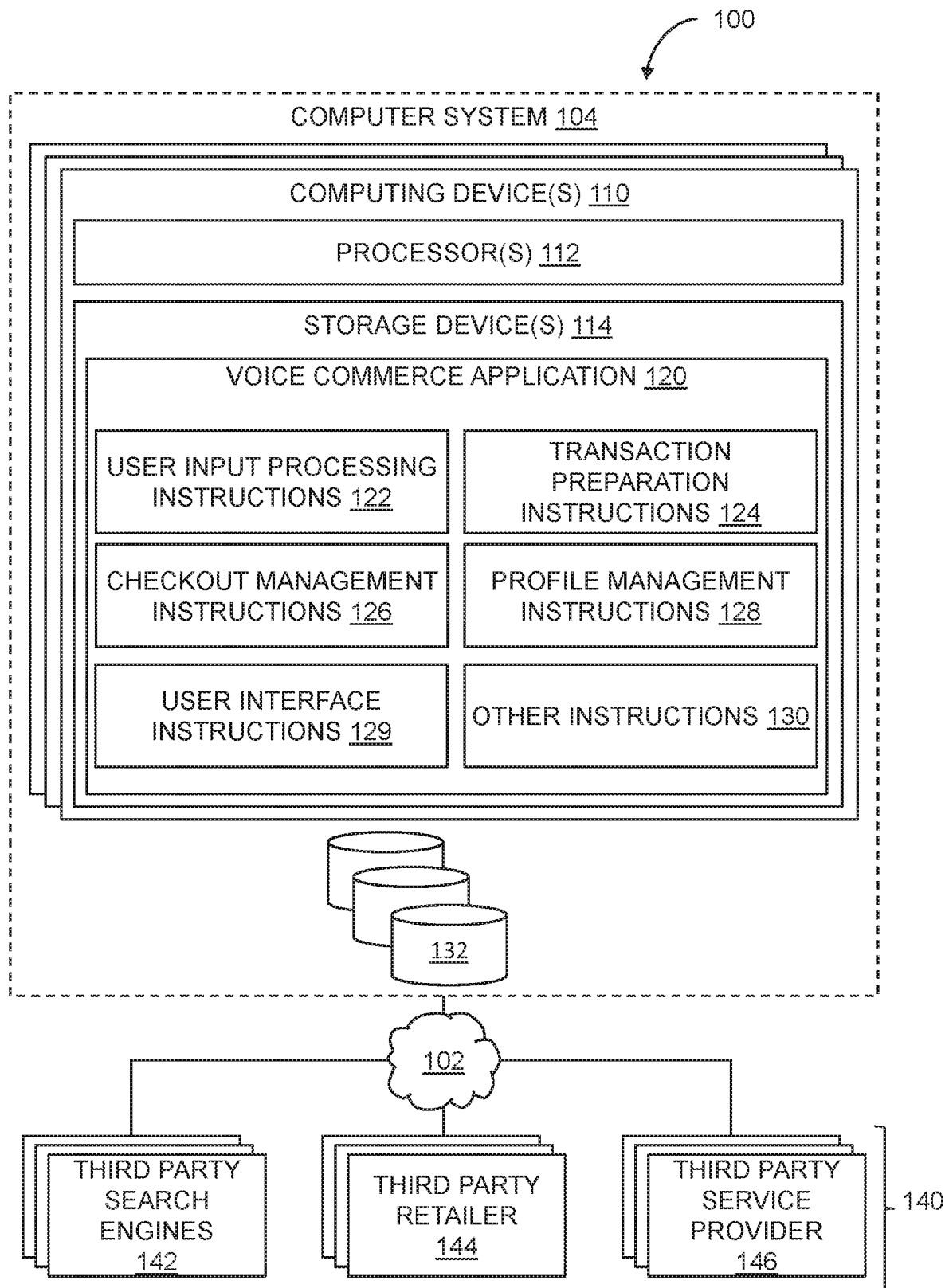


FIG. 1

U.S. Patent

Aug. 10, 2021

Sheet 2 of 6

US 11,087,385 B2

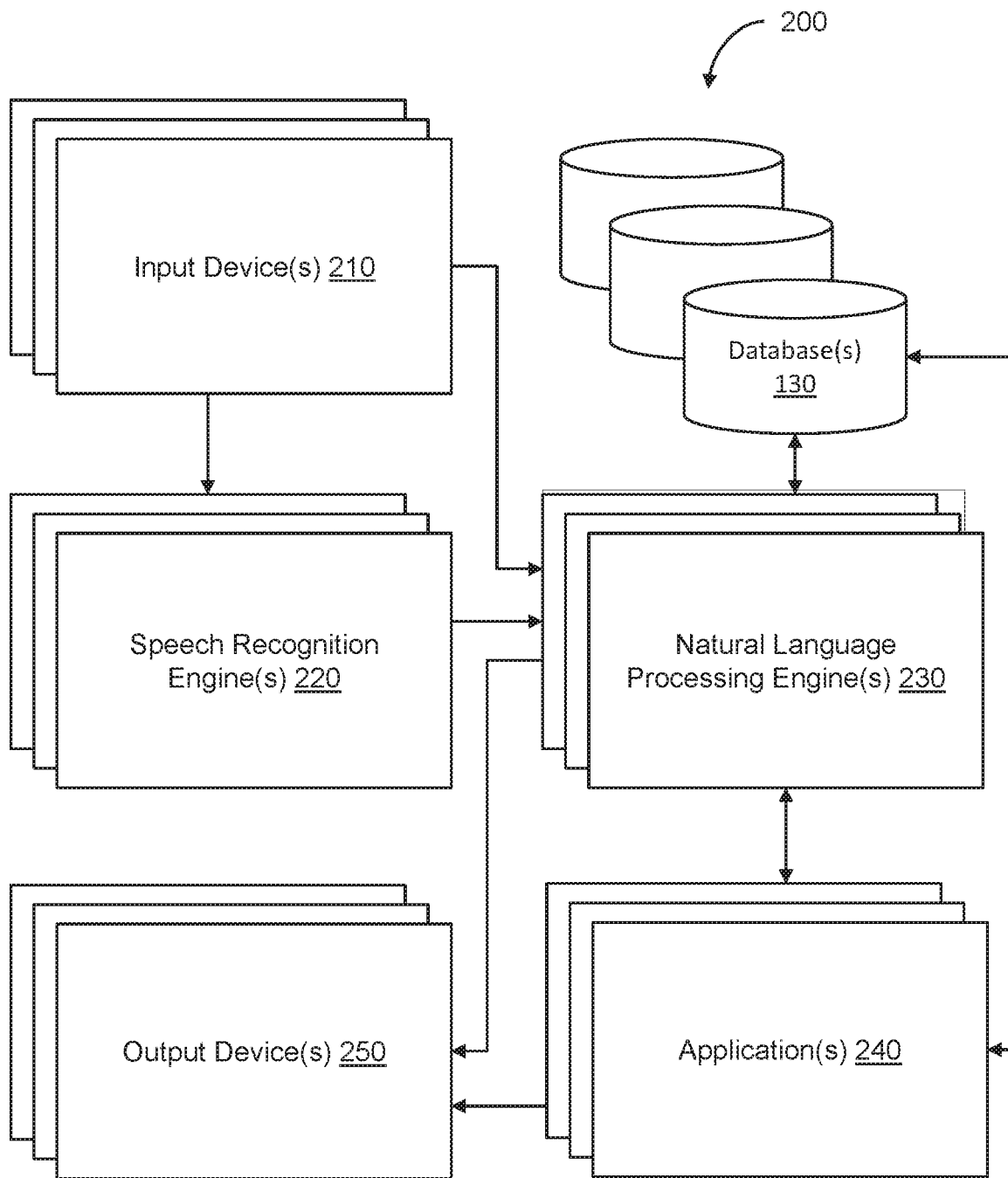


FIG. 2

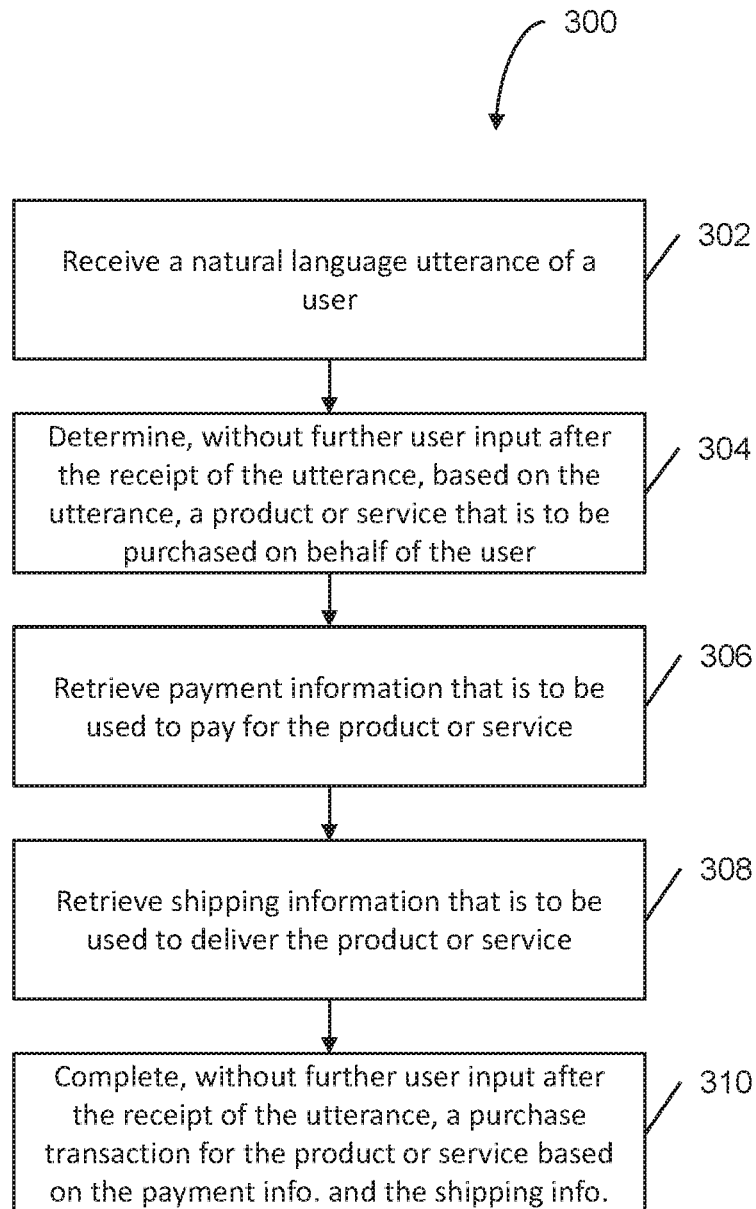


FIG. 3

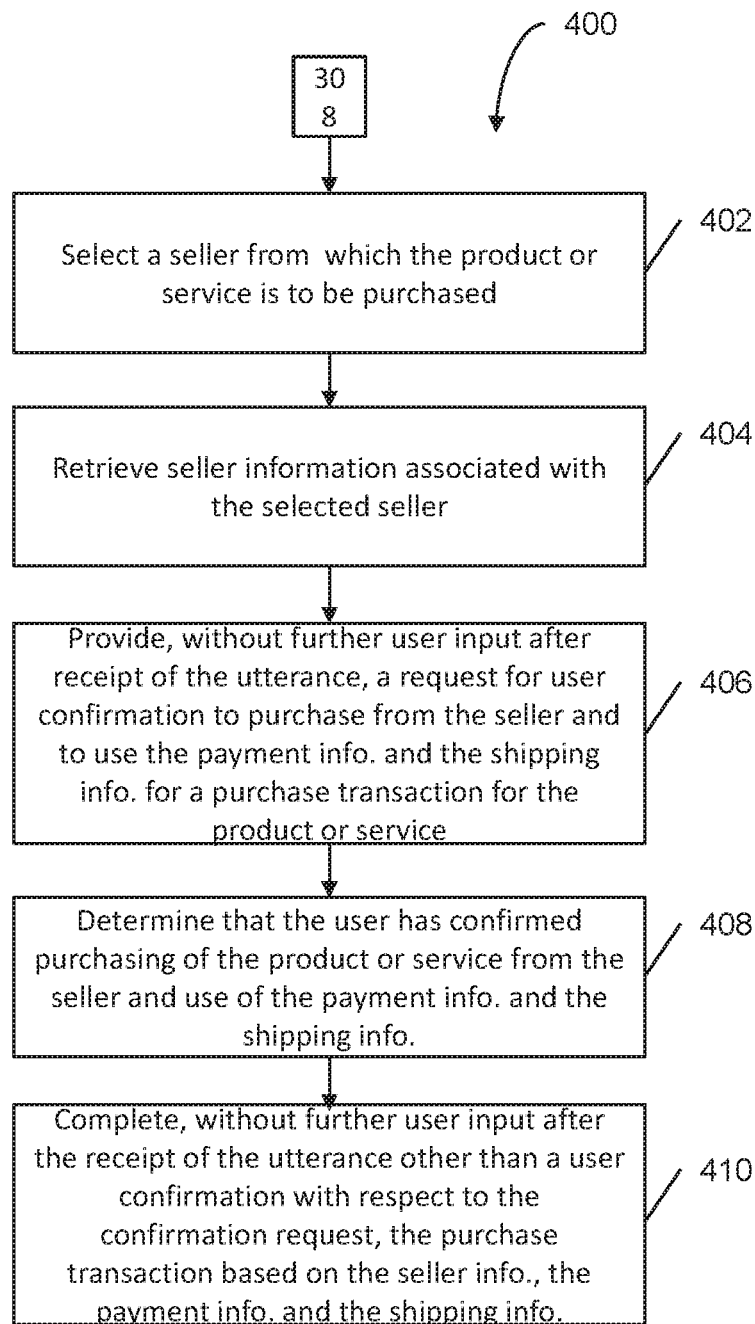


FIG. 4

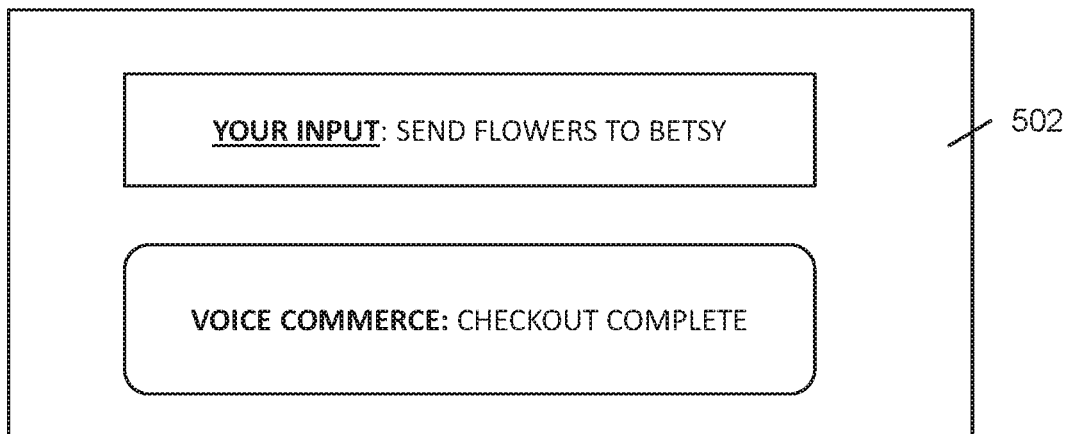


FIG. 5



FIG. 6A

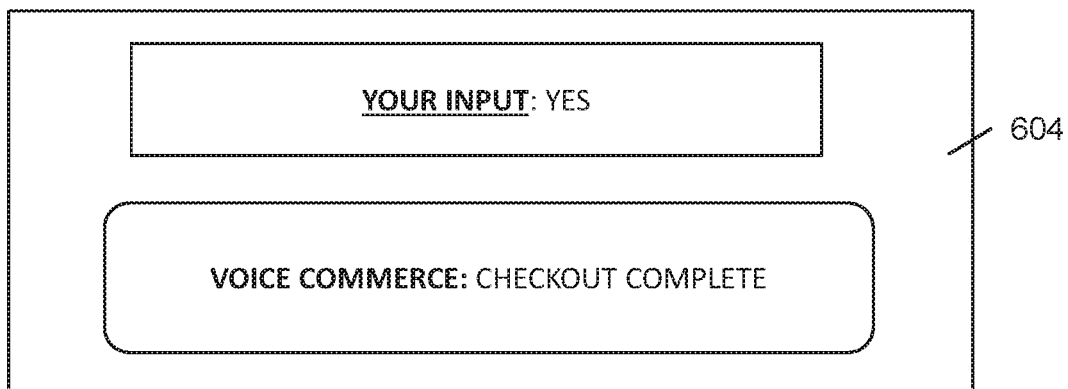


FIG. 6B

U.S. Patent

Aug. 10, 2021

Sheet 6 of 6

US 11,087,385 B2

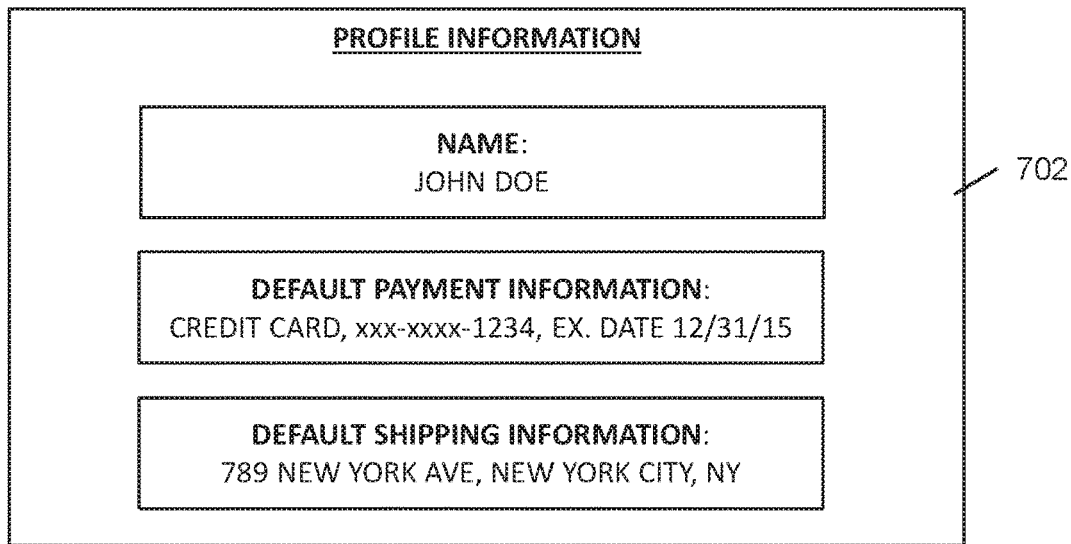


FIG. 7A is a diagram of a user profile information form, designated by reference numeral 702. The form is titled "PROFILE INFORMATION" and contains three stacked input fields. The first field is labeled "NAME:" and contains the text "JOHN DOE". The second field is labeled "DEFAULT PAYMENT INFORMATION:" and contains the text "CREDIT CARD, xxx-xxxx-1234, EX. DATE 12/31/15". The third field is labeled "DEFAULT SHIPPING INFORMATION:" and contains the text "789 NEW YORK AVE, NEW YORK CITY, NY".

FIG. 7A

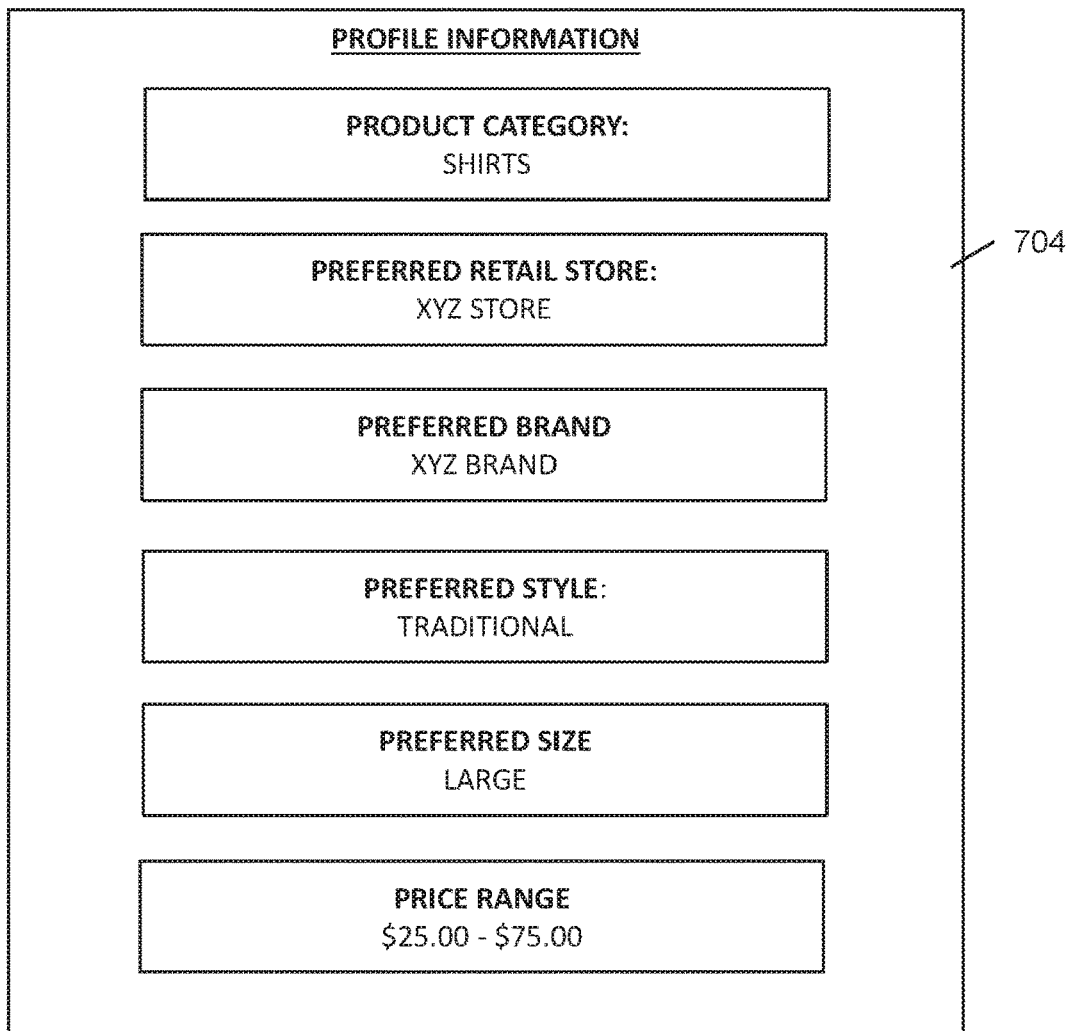


FIG. 7B is a diagram of a user profile information form, designated by reference numeral 704. The form is titled "PROFILE INFORMATION" and contains six stacked input fields. The first field is labeled "PRODUCT CATEGORY:" and contains the text "SHIRTS". The second field is labeled "PREFERRED RETAIL STORE:" and contains the text "XYZ STORE". The third field is labeled "PREFERRED BRAND" and contains the text "XYZ BRAND". The fourth field is labeled "PREFERRED STYLE:" and contains the text "TRADITIONAL". The fifth field is labeled "PREFERRED SIZE" and contains the text "LARGE". The sixth field is labeled "PRICE RANGE" and contains the text "\$25.00 - \$75.00".

FIG. 7B

US 11,087,385 B2

1

VOICE COMMERCE

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. patent application Ser. No. 15/484,930, filed Apr. 11, 2017, entitled “VOICE COMMERCE”, which is a continuation of U.S. patent application Ser. No. 14/855,334, filed Sep. 15, 2015, entitled “VOICE COMMERCE” (now U.S. Pat. No. 9,626,703 issued Apr. 18, 2017), which claims the benefit of U.S. Provisional Patent Application Ser. No. 62/051,273, filed Sep. 16, 2014, entitled “VOICE COMMERCE”, all of which are hereby incorporated by reference in their entireties.

FIELD OF THE INVENTION

The invention relates to systems and methods for providing voice commerce. More particularly, the invention relates to systems and methods for preparing and/or completing checkout of product or service purchases via a single utterance.

BACKGROUND OF THE INVENTION

Mobile electronic devices have emerged to become nearly ubiquitous in the everyday lives of many people. One of the reasons for this increased use is the convenience of performing tasks with a mobile electronic device. One task that has seen significant growth is online shopping. During an online shopping session, a user browses a website to locate a product or service for purchase. After the product or service has been located, the user makes payment through the mobile electronic device and has the purchased product or service delivered.

One difficulty in online shopping via a mobile electronic device is that the user must search a website in order to locate a product or service to be purchased and fill out numerous payment and shipping forms to complete checkout (or a purchase transaction). Mobile electronic devices typically contain small screens and keyboards, making it hard for the user to search for the product or service to purchase and input payment and shipping information. While some online shopping applications assist the user in filling out payment and shipping forms, the foregoing applications are often limited to the amount of information that can be provided. Further, some online shopping applications include a “one-click” purchase option. However, this still requires the user to browse the website to locate product or services which they wish to purchase. These and other drawbacks exist.

SUMMARY OF THE INVENTION

The invention relates to systems and methods for preparing and/or completing checkout of product or service purchases via a single utterance. More particularly, the systems and methods may process a single utterance of a user to determine a product or service that is to be purchased, payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, and/or other information that is thereafter utilized to prepare and/or complete a purchase transaction for the product or service. In some implementations, the preparation and/or completion of the purchase transaction may be performed without further user input

2

after receipt of the utterance (with which the determination of the product or service, the payment information, or shipping information is based).

In an implementation, the system may receive and process a user input comprising a natural language utterance to determine a product or service to be purchased on behalf of a user. As an example, one or more words associated with the natural language utterance may be indicative of a product or service type, product name, seller name, etc., which can be used to determine the product or service that is to be purchased. Without further user input after the receipt of the user input (or the natural language utterance), the system may obtain payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, or other information that is to be used to complete a purchase transaction for the product or service.

In an implementation, upon receipt of a natural language utterance of a user, and without further user input after the receipt of the utterance, the system may process the utterance to determine a product or service, and complete a purchase transaction for the product or service based on payment information associated with the user, shipping information indicating who, where, when, and/or how the product or service is to be delivered, or other information. In a further implementation, after the purchase transaction is completed, the user may be provided with an option to modify one or more aspects related to the product or service purchase (e.g., modify payment information, shipping information, seller from which the product or service is purchased, etc.). As an example, the initial product or service purchase may be cancelled (and result in refund of the payment provided for the completed purchase transaction) in response to the user’s modification, and a new purchase transaction may be initiated in accordance with the user’s modification. Of course, in some implementations, such after-transaction-completion modifications may be limited to avoid fraud or other issues (e.g., limited to a predefined time period after the transaction completion, limited to certain types of modifications, etc.). In this way, users can experience the convenience of a purchase transaction being completed with reduced actions (e.g., single utterance) without fear of inaccuracies that might be associated with the purchase transaction.

In an implementation, upon receipt of a natural language utterance of a user, and without further user input after the receipt of the utterance, the system may process the utterance to determine a product or service to be purchased, and present a request for user confirmation (or user approval) of the product or service to be purchased, payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, or other information that is to be used to complete a purchase transaction for the product or service. Upon receipt of the requested confirmation from the user, a purchase transaction for the product or service may be completed without further user input after the receipt of the requested confirmation. As an example, after the user has provided the utterance, the user may be presented with a prompt that: (i) identifies the product or service, the cost(s) associated with the purchase/delivery of the product or service, the payment information, and the shipping information; and (ii) solicits the user’s approval to complete checkout of the product or service using a particular payment method (specified by the payment information to pay for the associated cost(s)), a particular shipping method (specified by the shipping information), along with a name

US 11,087,385 B2

3

and address (specified by the shipping information) to which the product or service is to be delivered. The user may respond by providing a subsequent natural language utterance. If it is determined (upon receiving and processing the subsequent utterance) that the user provided his/her approval, checkout of the product or service purchase may be completed (e.g., without further user input after receipt of the utterance other than the subsequent utterance). In this way, the confirmation may be used to ensure that the information utilized to complete checkout of the product or service checkout is accurate and acceptable to the user.

Various other aspects of the invention will be apparent through the detailed description of the invention and the drawings attached hereto. It is also to be understood that both the foregoing general description and the following detailed description are exemplary and not restrictive of the scope of the invention. As used in the specification and in the claims, the singular form of “a”, “an”, and “the” include plural referents unless the context clearly dictates otherwise. In addition, as used in the specification and the claims, the term “or” means “and/or” unless the context clearly dictates otherwise.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a system for providing voice commerce, according to an implementation of the invention.

FIG. 2 illustrates a system for facilitating natural language processing, according to an implementation of the invention.

FIG. 3 illustrates a flow diagram of a process for providing voice commerce, according to an implementation of the invention.

FIG. 4 illustrates a flow diagram of another process for providing voice commerce, according to an implementation of the invention.

FIG. 5 illustrates a screenshot of a user interface which facilitates voice commerce, according to an implementation of the invention.

FIGS. 6A and 6B illustrate screenshots of a user interface which facilitates voice commerce, according to an implementation of the invention.

FIGS. 7A and 7B illustrate screenshots of a user interface which facilitates management of profile information, according to an implementation of the invention.

DETAILED DESCRIPTION OF THE INVENTION

In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the implementations of the invention. It will be appreciated, however, by those having skill in the art that the implementations of the invention may be practiced without these specific details or with an equivalent arrangement. In other instances, well-known structures and devices are shown in block diagram form in order to avoid unnecessarily obscuring the implementations of the invention.

FIG. 1 illustrates a system **100** for providing voice commerce to a user, according to an implementation of the invention. In an implementation, the system **100** may prepare and/or complete checkout of product or service purchases related to one or more user inputs. The user inputs may comprise an auditory input (e.g., received via a microphone), a visual input (e.g., received via a camera), a tactile input (e.g., received via a touch sensor device), a keyboard input, a mouse input, or other user input that is recognizable

4

by the system **100**. As an example, a natural language utterance may be received from a user and processed to prepare and/or complete checkout of product or service purchases.

In an implementation, system **100** may receive and process a single utterance of a user to determine a product or service that is to be purchased, payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, and/or other information that is thereafter utilized to prepare and/or complete the purchase of the product or service.

In one use case, if the user input is a natural language utterance spoken by a user, the natural language utterance may be processed by a speech recognition engine to recognize one or more words of the natural language utterance. The recognized words may then be processed, along with context information associated with the user, by a natural language processing engine to determine one or more products or services and/or other information (e.g., payment information, shipping information, seller information, associated cost information, etc.) that can be used to prepare and/or complete one or more purchase transactions for the products or services. Upon determination of such information, for example, the natural language processing engine may provide the information to one or more applications that then utilize the information to prepare and/or complete the purchase transactions.

As an example, when a user speaks the utterance “Send flowers to my mother,” the user’s mother may be identified as the recipient of a product that the user intends to purchase (e.g., based on the words “send,” “my,” and “mother”), and flowers may be identified as the type of product that the user intends to have purchased and delivered to the user’s mother (e.g., based on the words “send” and “flowers”). Without necessarily receiving any further input from the user, the system **100** may automatically determine the name and address of the user’s mother, a payment method associated with the user, a seller from which flowers are to be purchased, a particular item (e.g., a particular flower purchase order) to be purchased from the seller, and the costs associated with the purchase/delivery of the purchase item. Such information may then be utilized to initiate and/or complete a purchase transaction for the purchase item. The name and address of the user’s mother may, for example, be determined by querying the user’s profile information, the user’s address book(s) or contact list(s), the user’s social network(s), one or more people search databases, etc. The payment method associated with the user may be determined by querying the user’s profile information. The seller from which flowers are to be purchased or the purchase item from the seller may be determined based on a set of preferred merchants indicated by a service provider (e.g., a company associated with the system **100**), preference information associated with the user that identifies merchants approved by the user (e.g., the user’s profile information may identify a predefined set of merchants from which the user has approved the system **100** to purchase products or services on behalf of the user), or similar criteria. For example, system **100** may determine which flowers are to be purchased from the merchant’s range of offerings based on the user’s profile information including the types of items a user likes, qualities of items the user likes, the categories of items a user likes, minimum item rating requirements, item pricing requirements, user location and context information, shipping requirements, etc. The associated costs may be determined based on information regarding the seller, the pur-

US 11,087,385 B2

5

chase item, the mother's address, the delivery method (e.g., different fees associated with one-day shipping, two-day shipping, etc.), the payment method (e.g., different fees associated with different payment methods), or other information.

It should be noted that while, in some implementations, a purchase transaction for a product or service may be completed without further user input after receiving a natural language utterance (from which the product or service is determined), user input may be solicited (or received unsolicited) from a user in other implementations to determine payment information, shipping information, seller information, or other information that is to be used to complete the purchase transaction.

In an implementation, upon receipt of a natural language utterance of a user, and without further user input after the receipt of the utterance, the system 100 may process the utterance to determine a product or service to be purchased, and present (e.g., via a graphical user interface, via an auditory user interface, etc.) a request for user confirmation (or user approval) of the product or service, shipping information that is to be used to deliver the product or service, payment information that is to be used to pay for the product or service, and/or other information to ensure that such information is correct and accepted by the user for use with purchasing the product or service. Upon receipt of the requested confirmation from the user, a purchase transaction for the product or service may be completed without further user input after the receipt of the requested confirmation.

As an example, with respect to the above "Send flowers to my mother" utterance scenario, the following information may be presented to the user as part of the user confirmation request: (i) the name and address of the user's mother; (ii) a payment method associated with the user; (iii) a seller from which flowers are to be purchased; (iv) a particular item (e.g., a particular flower purchase order) to be purchased from the seller; and (v) the costs associated with the purchase/delivery of the purchase item. As a further example, the user may provide any of the following utterances—"That's fine," "The information is correct," "Go ahead and purchase the flowers," "Order the flowers," or other utterance—to indicate the user's confirmation to complete the purchase transaction. As such, based on the confirmation, the purchase transaction for the purchase item (e.g., a particular flower purchase order) may be completed (e.g., without further user input after the receipt of the utterance "Send flowers to my mother" other than the confirmation utterance).

In an implementation, user profile information (e.g., name, payment information, shipping information, preferences, etc.) may be pre-stored so that the user profile information may be utilized by system 100 for preparing and/or completing checkout of product or service purchases. In an implementation, defaults associated with the user profile information, such as default payment information, default shipping information, etc., may be automatically or manually pre-set for preparing and/or completing checkout of product or service purchases.

In an implementation, preferred sellers, brands, style, size, or other parameters related to products or services may be automatically or manually pre-set as profile information. As an example, preferred sellers from which products or services are to be purchased and preferred brands of such products or services may be automatically pre-set by the system (e.g., pre-set to system preferred sellers and brands) until the preferences are manually modified by the user. The preferred store, brand, style, size, or other parameters may

6

be obtained (e.g., in response to a natural language utterance indicating a product or service without further user input after the receipt of the utterance) to prepare and/or complete checkout of a product or service purchase unless the utterance (or other information) specifies otherwise. In another implementation, the profile information may include information relating to the general likes or dislikes of the user utilized to select the product or service to be purchased including the types of items a user likes, qualities of items the user likes, the categories of items a user likes, minimum item rating requirements, item pricing requirements, user location and context information, shipping requirements, etc.

In an implementation, preparation and/or completion of checkout of product or service purchases may be limited to certain categories of products or services (e.g., no automobiles), a particular price range (e.g., no more than \$100), etc., to reduce the number or severity of purchases related to fraud. The categories of product or services and/or the price ranges may, for instance, be automatically or manually pre-set (e.g., automatically preset by the system, manually pre-set by the user).

Other uses of system 100 are described herein and still others will be apparent to those having skill in the art. Having described a high level overview of some of the system functions, attention will now be turned to various system components that facilitate these and other functions.

System Components

System 100 may include a computer system 104, one or more databases 132, one or more remote information sources 142, 144, 164, and/or other components.

To facilitate these and other functions, computer system 104 may include one or more computing devices 110. Each computing device 110 may include one or more processors 112, one or more storage devices 114, and/or other components.

Processor(s) 112 may be programmed by one or more computer program instructions, which may be stored in storage device(s) 114. The one or more computer program instructions may include, without limitation, voice commerce application 120. Voice commerce application 120 may itself include different sets of instructions that each program the processor(s) 112 (and therefore computer system 104) to perform one or more operations described herein. For example, voice commerce application 120 may include user input processing instructions 122, transaction preparation instructions 124, checkout management instructions 126, profile management instructions 128, user interface instructions 129, and/or other instructions 130 that program computer system 104. Other applications may, of course, include one or more of the instructions 120-130 to perform one or more operations as described herein. As used herein, for convenience, the various instructions will be described as performing an operation, when, in fact, the various instructions program computer system 104 to perform the operation.

In some implementations, a given user device 160 may comprise a given computer device 110. As such, the given user device 160 may comprise processor(s) 112 that are programmed with one or more computer program instructions, such as voice commerce instructions 120, user input processing instructions 122, transaction preparation instructions 124, checkout management instructions 126, profile management instructions 128, user interface instructions 129, and/or other instructions 130.

As used hereinafter, for convenience, the foregoing instructions will be described as performing an operation,

US 11,087,385 B2

7

when, in fact, the various instructions may program processor(s) **112** (and thereafter computer system **104**) to perform the operation.

Registering User Information and Providing Voice Commerce

In an implementation, voice commerce application **120** may register a user to use the system. For example, during registration of a user, voice commerce application **120** may obtain profile information of the user that includes user settings. Information obtained during registration (or registration information) may include, for example, user identification information, payment information, shipping information, user preferences, and/or other information. The registration information may also include preferred sellers, brands, style, size, or other parameters related to products or services as well as limits on the purchase of particular product and service categories and particular price ranges for products or services. In another implementation, the registration information may also include limitations of certain categories of products or services to be purchased and/or particular price ranges of the products or services to be purchased. The registration information may be stored as profile information associated with the user in one or more databases, such as a database **132**.

In an implementation, the voice commerce application **120** may process one or more user inputs to prepare and/or complete checkout of product or service purchases related to the one or more user inputs. The user inputs may comprise an auditory input (e.g., received via a microphone), a visual input (e.g., received via a camera), a tactile input (e.g., received via a touch sensor device), an olfactory input, a gustatory input, a keyboard input, a mouse input, or other user input. As an example, a natural language utterance, a natural language gesture, or other natural language user input may be received from a user, and processed to determine the meaning of the user input (and corresponding actions, if any, to be taken) with respect to preparation and completion of a purchase transaction. It should be noted that while, in some implementations, the receipt of a natural language utterance may trigger preparation and/or completion of a purchase transaction with further user input after the receipt of the utterance, the receipt of a natural language input of another type (e.g., gestures or other non-verbal communication) may, in other implementations, trigger preparation and/or completion of a purchase transaction with further user input after the receipt of the input of the other type. As such, to the extent possible, one or more operations described herein as based on a natural language utterance may, in other implementations, instead be based on a natural language input of another type.

The voice commerce application **120** may utilize instructions associated with one or more speech recognition engines, one or more natural language processing engines, or other components for processing user inputs to determine user requests related to the user inputs. For example, voice commerce application **120** may process a single utterance of a user to determine a product or service that is to be purchased, payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, and/or other information that is thereafter utilized to prepare and/or complete the purchase of the product or service.

In an implementation, voice commerce application **120** may process a user input (e.g., a natural language utterance) to determine one or more words associated with the user input, and initiate one or more user requests (e.g., queries, commands, etc.) based on the determined words, context

8

information associated with the user (e.g., a current location of the user, the time at which the user spoke the utterance, a product or service recently discussed by the user, or other context), user profile information, or other information. As an example, a user may speak the utterance "Order a pizza for delivery," and the voice commerce application **120** may interpret the utterance and determine that the user's intent is to purchase a pizza for delivery to the user. In one use case, without further user input after receiving the utterance, the voice commerce application **120** may initiate and complete a purchase transaction for the pizza to be delivered. If, for instance, the voice commerce application **120** determines that the user is currently at home, it may complete a purchase transaction with a pizza restaurant from which the user has previously ordered pizza to have the previously ordered pizza delivered to the user's house.

In an implementation, voice commerce application **120** may utilize a user input (e.g., a natural language utterance) to determine a product or service that is to be purchased by searching one or more databases associated with one or more sellers (e.g., partners, third party retailers, service providers, etc.). Results obtained from the search may include one or more products and/or services related to the user input. Each set of results may be individually grouped according to its source or product/service category. The voice commerce application **120** may automatically select the product/service to be purchased from the set of results. As an example, a single utterance may automatically result in the selection of a seller (e.g., a third party retailer) and the selection of a product related to the utterance that is available for purchase via the seller.

In an implementation, the voice commerce application **120** may prepare and/or complete checkout of a product or service purchase related to one or more user inputs. The voice commerce application **120** may determine payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, and/or other information that is thereafter utilized to prepare and/or complete the purchase of the product or service. In one implementation, the voice commerce application **120** may automatically complete the checkout of the product or service purchase related to the one or more user inputs. As an example, a single utterance from the user may result in the preparation and/or completion of the checkout of the product or service without further input from the user. In another implementation, the voice commerce application **120** may provide a request for confirmation of the determined product or service, shipping information, payment information, and/or other information to the user to solicit a confirmation from the user to complete the checkout of the product or service purchase.

In an implementation, the voice commerce application **120** may utilize user profile information to prepare and/or complete checkout of product or service purchases via a user input. In an implementation, the voice commerce application **120** may utilize defaults associated with the user profile information, such as default payment information, default shipping information, etc., for preparing and/or completing checkout of product or service purchases. In another implementation, the voice commerce application **120** may utilize product and/or service preferences including preferred sellers, brands, style, size, or other parameters related to products or services for preparing and/or completing checkout of product or service purchases. In another embodiment, the voice commerce application **120** may limit checkout of product or service purchases to certain categories of products or services, a particular price range, etc., to reduce the

US 11,087,385 B2

9

number or severity of purchases related to fraud. In another implementation, the profile information may determine a product or service to be purchased based from profile information relating to the general likes or dislikes of the user including the types of items a user likes, qualities of items the user likes, the categories of items a user likes, minimum item rating requirements, item pricing requirements, user location and context information, shipping requirements, etc.

Having described high level functions and operations of voice commerce application 120, attention will now be turned to particular functions and operations of voice commerce application 120 as illustrated through its various instructions. The various instructions (e.g., user input processing instructions 122, transaction preparation instructions 124, checkout management instructions 126, profile management instructions 128, user interface instructions 129, and/or other instructions 130) of voice commerce application 120 are described individually as discreet sets of instructions by way of illustration and not limitation, as two or more of the instructions may be combined.

User Input Processing

In an implementation, the user input processing instructions 122 may process one or more user inputs of a user to prepare and/or complete checkout of product or service purchases related to the one or more user inputs. The user inputs may comprise an auditory input (e.g., received via a microphone), a visual input (e.g., received via a camera), a tactile input (e.g., received via a touch sensor device), a keyboard input, a mouse input, or other user input. As described herein elsewhere, user input processing instructions 122 may comprise instructions associated with one or more speech recognition engines (e.g., speech recognition engine(s) 220 of FIG. 2), one or more natural language processing engines (e.g., natural language processing engine(s) 230 of FIG. 2), or other components for processing user inputs to determine user requests related to the user inputs. For example, user input processing instructions 122 may process a natural language utterance of a user to determine a product or service that is to be purchased, payment information that is to be used to pay for the product or service, shipping information that is to be used to deliver the product or service, and/or other information that is thereafter utilized to prepare and/or complete the purchase of the product or service.

In one use case, if the user input is a natural language utterance spoken by a user, the natural language utterance may be processed by a speech recognition engine to recognize one or more words of the natural language utterance. The recognized words may then be processed, along with context information associated with the user, by a natural language processing engine to determine one or more products or services and/or other information (e.g., payment information, shipping information, seller information, associated cost information, etc.) that can be used to prepare and/or complete checkout of product or service purchases when the user provided the natural language utterance.

In an implementation, the user input processing instructions 122 may utilize one or more previous user inputs (e.g., related to a product or service) in processing a particular user input to determine one or more products or services and/or other information (e.g., payment information, shipping information, seller information, associated cost information, etc.) that can be used to prepare and/or complete one or more checkout of product or service purchases.

As an example, a first user input (that occurs prior to a second user input) may be indicative of a user's intent to

10

make a purchase (e.g., without necessarily identifying a product or service type or a product or service), and the second user input may be indicative of a particular product or service type and/or a particular product or service (e.g., without necessarily being indicative of the user's intent to make a purchase). In one use case, for example, a user may provide the utterance "I'm looking to buy something" as the first user input, after which an automated personal assistant (of a related application) may ask the user what the user would like to buy. The user may respond with the second user input by saying "Flowers for my mom." The word "buy" may be recognized when speech recognition is performed on the first user input, and the word "flowers" may be recognized when speech recognition is performed on the second user input. Further processing on the word "buy" may indicate the user's intent to purchase, and further processing on the word "flowers" may indicate the particular product type of a product that the user would like to purchase.

As another example, a first user input (that occurs prior to a second user input) may be indicative of a particular product or service type and/or a particular product or service (e.g., without necessarily being indicative of the user's intent to make a purchase), and the second user input may be indicative of a user's intent to make a purchase (e.g., without necessarily identifying a product or service type or a product or service). In one scenario, for example, the user may provide the utterance "Those are beautiful flowers" while viewing photographs on a friend's social networking page. Subsequently, the user may provide the utterance "I'd like to buy some." The words "those" and "flowers" may be recognized when speech recognition is performed on the first user input, and the words "buy" and "some" may be recognized when speech recognition is performed on the second user input. Further processing on the words "those" and "flowers" may indicate the flowers in the photographs that the user viewed on the friend's social networking page, and further processing on the words "buy" and "some" may indicate the user's intent to make a purchase.

FIG. 2 illustrates a system 200 for facilitating natural language processing, according to an implementation of the invention. As shown in FIG. 2, system 200 may comprise input device(s) 210, speech recognition engine(s) 220, natural language processing engine(s) 230, application(s) 240, output device(s) 250, database(s) 132, or other components.

In an implementation, one or more components of system 200 may comprise one or more computer program instructions of FIG. 1 and/or processor(s) 112 programmed with the computer program instructions of FIG. 1. As an example, speech recognition engine(s) 220 and/or natural language processing engine(s) 230 may comprise user input processing instructions 122, transaction preparation instructions 124, checkout management instructions 126, profile management instructions 128, or other instructions.

Input device(s) 210 may comprise an auditory input device (e.g., microphone), a keyboard, a mouse, or other input devices. Input received at input device(s) 210 may be provided to speech recognition engine(s) 220 and/or natural language processing engine(s) 230.

Speech recognition engine(s) 220 may process one or more inputs received from input device(s) 210 to recognize one or more words represented by the received inputs. As an example, with respect to auditory input, speech recognition engine(s) 220 may process an audio stream captured by an auditory input device to isolate segments of sound of the audio stream. The sound segments (or a representation of the sound segments) are then processed with one or more speech

US 11,087,385 B2

11

models (e.g., acoustic model, lexicon list, language model, etc.) to recognize one or more words of the received inputs. Upon recognition of the words of received inputs, the recognized words may then be provided to natural language processing engine(s) **230** for further processing. In other examples, natural language processing engine(s) **230** may process one or more other types of inputs (e.g., visual input representing sign language communication, gestures, or other forms of communication) to recognize one or more words represented by the other types of inputs.

Natural language processing engine(s) **230** may receive one or more inputs from input device(s) **210**, speech recognition engine(s) **220**, application(s) **240**, database(s) **132**, or other components. As an example, natural language processing engine(s) **230** may process inputs received from input device(s) **210**, such as user inputs (e.g., voice, non-voice, etc.), location-based inputs (e.g., GPS data, cell ID, etc.), other sensor data input, or other inputs to determine context information associated with one or more user inputs. As another example, natural language processing engine(s) **230** may obtain profile information, context information, or other information from database(s) **132**. The obtained information (or context information determined based on inputs from input device(s) **210**) may be processed to determine one or more user inputs of a user. In yet another example, natural language processing engine(s) **230** may process one or more recognized words from speech recognition engine(s) **220** and other information (e.g., information from input device(s) **210**, application(s) **240**, and/or database(s) **132**) to determine one or more user inputs.

In an implementation, natural language processing engine(s) **230**, application(s) **240**, or other components may store information in database(s) **132** for later use by natural language processing engine(s) **230**, application(s) **240**, or other components. As an example, as described in further detail elsewhere herein, natural language processing engine(s) **230** may store information regarding user inputs in database(s) **132** and/or update profile information, or other information in database(s) **132** based on the information regarding the user inputs.

Transaction Preparation and Search for Products or Services

In an implementation, transaction preparation instructions **124** may utilize information from a processing of the user inputs (e.g., one or more recognized words, product or service type, product name, seller name, etc.) to determine a product or service that is to be purchased. Transaction preparation instructions **124** may utilize such information to search one or more databases associated with one or more sellers (e.g., partners, third party retailers, service providers, etc.). Results obtained from the search may include one or more products and/or services related to the user input. The results are then utilized by the transaction preparation instructions **124** to select the product or service to be purchased by the user. As an example, a single utterance may automatically result in the selection of a seller (e.g., a third party retailer) and the selection of a product or service related to the utterance that is available for purchase via the seller.

In an implementation, transaction preparation instructions **124** may obtain product or service results related to the user input that are available from a remote information source **140**. Remote information sources **140** may include information sources that are accessible to computer system **104** via a remote or external network connection (e.g., outside of a firewall), such as the Internet. For example, as illustrated in FIG. 1, remote information source(s) **140** may include,

12

without limitation, a third party search engine **142** (e.g., GOOGLE, BING, YAHOO, etc.), third party retailers **144**, including online retailers and offline retailers having an online presence (e.g., AMAZON, GROUPON, EBAY, EXPEDIA, BESTBUY, etc.), third party service providers **146** (e.g., weather services, calendar/scheduling services, etc.), and/or other information source that is remotely accessible to computer system **104**. In another implementation, the transaction preparation instructions **124** may select the product or service (that is to be purchased) from one or more predetermined sets of products or services stored in database **132**. For example, system **200** may store one or more sets of predetermined products or services that may be selected by the transaction preparation instructions **124** or included in the product or service results.

In an implementation, transaction preparation instructions **124** may provide a set of product and service results for the search. The transaction preparation instructions **124** may utilize the set of results to select the product or service that is to be purchased. In an implementation, the transaction preparation instructions **124** may select the product or service (that is to be purchased) according to a set of predefined rules. For example, the product or service (that is to be purchased) may be selected based on the user input, location of the user, cost comparison of sellers (e.g., partners, third party retailers, third party service providers, etc.), shipping date comparison of the sellers, or other criteria. In another implementation, the results of the search are presented for selection by the user. In another implementation, the profile information may include information relating to the general likes or dislikes of the user to determine the product or service to be purchased including the types of items a user likes, qualities of items the user likes, the categories of items a user likes, minimum item rating requirements, item pricing requirements, user location and context information, shipping requirements, etc.

In an implementation, transaction preparation instructions **124** may utilize context information associated with the user to determine the product or service to be purchased. Based on the context information, the transaction preparation instructions **124** may refine the search for products or services to be purchased. In one embodiment, the context information may include personal data associated with the user, data from a database associated with the user, data describing an event, data describing an acoustic environment in which the spoken input is received, location, local time, etc. For example, the transaction preparation instructions **124** may utilize the time of the user input to further define the availability of product or service to be purchased.

Continuing the foregoing examples, a user input related to "lawnmower" may cause transaction preparation instructions **124** to cause a search on a third party search engine **142** to be initiated using the search term "lawnmower." In an implementation, depending on the context (e.g., indicating that the user intends to buy a lawnmower), transaction preparation instructions **124** may add additional search terms such as "purchase" or "sale." Transaction preparation instructions **124** may search (in addition to or instead of) other remote information sources **140** as well (e.g., retail related to "lawnmower" from one or more third party retailers **144**, landscaping services related to "lawnmower" from one or more third party service providers **146**, and/or other information related to "lawnmower" from one or more other remote information sources **140**).

In another embodiment, the transaction preparation instructions **124** may utilize user inputs such as location-based inputs (e.g., GPS data, cell ID, etc.) to further refine

US 11,087,385 B2

13

the search for products or services to be purchased. The transaction preparation instructions **124** may utilize the location-based input to further refine the geographic area in which the products or services are offered. For example, if the user utters "Please buy a pizza," the transaction preparation instructions **124** utilize the location-based inputs to determine the closest pizzeria in relation to the user. In an implementation, the transaction preparation instructions **124** may utilize website browsing information to refine the search for products or services to be purchased. For example, if the user utters "buy this" while browsing a retailing website, the transaction preparation instructions **124** may utilize website browsing information to determine the product or service the user is viewing. It should be contemplated that the transaction preparation instructions **124** prepares and/or completes checkout of the determined product or service directly from the website (or the seller associated with the website) or from another seller that may not be associated with the website.

In an implementation, the transaction preparation instructions **124** may utilize the profile information to select the product or service and/or the sellers from which the product or service is to be purchased. As described below, the profile information may include preferred sellers, brands, style, size, or other parameters related to products or services that may be automatically or manually pre-set. The profile information may also include the types of items a user likes, qualities of items the user likes, the categories of items a user likes, minimum item rating requirements, item pricing requirements, user location and context information, shipping requirements, etc. The transaction preparation instructions **124** may select the product or service and the sellers (from which the product or service is to be purchased) based on the profile information. As an example, preferred retailers from which products or services is to be purchased and preferred brands of such products or services may be automatically pre-set by the system (e.g., pre-set to system preferred sellers and brands) until the defaults are manually modified by users. The default store, brand, style, size, or other parameters may be obtained (in response to an utterance indicative of a user's intent to purchase a product or service) to prepare and/or complete checkout of a product or service purchase unless the single utterance (or other information) specifies otherwise.

In an implementation, selection of the product or service purchases by the transaction preparation instructions **124** may be limited to certain categories of products or services (e.g., no automobiles), a particular price range (e.g., no more than \$100), etc., to reduce the number or severity of purchases related to fraud. The categories of product or services and/or the price ranges may, for instance, be automatically or manually pre-set (e.g., automatically preset by the system, manually pre-set by the user) and stored in the profile information.

In one implementation, the transaction preparation instructions **124** may provide prospective transactions or offers to the user based on the user's profile information. For example, the transaction preparation instructions **124** may prepare a transaction for the user without any input from the user. In one implementation, the transaction preparation instructions **124** may analyze the user's profile information to determine any potential transaction opportunities, and provide a prospective transaction or offer related to the potential transaction opportunity. For example, in the case that voice commerce application **120** determines an upcoming anniversary of the user, the transaction preparation instructions **124** may prepare a prospective purchase for the

14

user related to the anniversary. In another implementation, the transaction preparation instructions **124** may provide offers to the user related to purchases for the potential transaction opportunity.

In an implementation, an administrator of the system **100** (or a subsystem thereof) may manage a set of predetermined sellers from whose inventories are searched for available products or services in response to a user input indicative of a user's intent to purchase. As an example, the administrator may specify a set of sellers that have priority over other sellers in having their products or services offered to users. Thus, a seller from which a product or service is to be purchased may be selected from the predetermined set of sellers specified by the administrator. Other criteria may, of course, be considered when selecting a seller with which a user is to complete a purchase transaction for a product or service.

Checkout Management

In an implementation, the checkout management instructions **126** may prepare and/or complete checkout of selected product or service purchases related to one or more user inputs. The checkout management instructions **126** may determine payment information that is to be used to pay for a selected product or service, shipping information that is to be used to deliver the selected product or service, and/or other information that is thereafter utilized to prepare and/or complete the purchase of the selected product or service. For example, the checkout management instructions **126** may utilize default payment information stored in the profile information for preparing or completing checkout of the selected product or service. Likewise, the checkout management instructions **126** may utilize default shipping information stored in the profile information for preparing or completing checkout of the selected product or service. In another implementation, the checkout management instructions **126** utilize an address book or contact list of the user stored in the profile information to provide shipping information for checkout of products or services for individuals other than the user. For example, a user input may reference an individual, other than the user, that is recognized by the user input processing instructions **122** (e.g., "Send flowers to my mom" wherein the user's mom is recognized as the intended recipient of the flowers). The checkout management instructions **126** may utilize the address book or contact list of the user to determine if the referenced individual's address is available for input as the shipping information.

In one implementation, the checkout management instructions **126** automatically complete the checkout of the product or service purchase. As an example, a single utterance from the user may result in the completion of checkout of the product or service without further input or approval from the user. In another implementation, the checkout management instructions **126** may present a request for confirmation of the determined product or service, shipping information, payment information, and/or other information to the user to solicit a confirmation from the user to complete the checkout of the product or service purchase. For example, the checkout management instructions **126** may prepare all of the aspects (e.g., payment information, shipping information, etc.) of checkout of the product or service purchase but wait until confirmation from the user to complete checkout.

In an implementation, the checkout management instructions **126** may utilize user profile information for preparing and/or completing checkout of product or service purchases via a user input. In an implementation, the checkout management instructions **126** may utilize defaults associated

US 11,087,385 B2

15

with the user profile information, such as default payment information, default shipping information, etc., for preparing and/or completing checkout of product or service purchases. In another implementation, the checkout management instructions 126 utilize an address book or contact list of the user stored in the profile information to provide shipping information for checkout of products or services for individuals other than the user.

Profile Management

In an implementation, stored user profile information (e.g., name, payment information, shipping information, preferences, etc.) may be utilized by voice commerce application 120 for preparing and/or completing checkout of product or service purchases. In an implementation, profile management instructions 128 may automatically set defaults associated with the user profile information, such as default payment information, default shipping information, etc. In another implementation, profile management instructions 128 may automatically or manually pre-set profile information for preparing and/or completing checkout of product or service purchases. In an implementation, profile management instructions 128 enable the user to set profile information including default payment and shipping information utilized by the voice commerce application 120.

In an implementation, profile management instructions 128 may automatically or pre-set preferred sellers, brands, style, size, or other parameters related to products or services. As an example, preferred retailers or service providers from which products or services are to be purchased and preferred brands of such products or services may be automatically pre-set by the profile management instructions 128 (e.g., pre-set to system preferred sellers and brands) until the preferences are manually modified by users. The preferred store, brand, style, size, or other parameters may be obtained (in response to a single utterance) to prepare and/or complete checkout of a product or service purchase unless the single utterance (or other information) specifies otherwise. In an implementation, profile management instructions 128 enable the user to set preferred sellers, brands, style, size, or other parameters related to products or services utilized by the voice commerce application 120. The profile management instructions 128 may automatically or pre-set information relating to the general likes or dislikes of the user to determine a product or service to be purchased including the types of items a user likes, qualities of items the user likes, the categories of items a user likes, minimum item rating requirements, item pricing requirements, user location and context information, shipping requirements, etc.

In an implementation, profile management instructions 128 may automatically or pre-set limits of checkout of product or service purchases to certain categories of products or services (e.g., no automobiles), a particular price range (e.g., no more than \$100), etc., to reduce the number or severity of purchases related to fraud. The categories of product or services and/or the price ranges may, for instance, be automatically or manually pre-set (e.g., automatically preset by the system, manually pre-set by the user). In an implementation, profile management instructions 128 enable the user to set limits of checkout of product or service purchases to certain categories of products or services (e.g., no automobiles), a particular price range (e.g., no more than \$100), or other criteria.

In another implementation, profile management instructions 128 may store profile information relating to the user's personal information. For example, the profile information may include the user's address book or contact list, calendar, and other information which assist the voice commerce

16

application 120 in determining products or services to be purchased. In one implementation, the user's personal information may be pre-stored based on the registration information or otherwise set by the user.

User Interface

In an implementation, user interface instructions 129 may generate a voice commerce interface. The voice commerce interface may provide status information relating to the preparation and/or completion of checkout of product or service purchases. For example, the voice commerce interface may indicate to a user when checkout of product or service purchases is completed. In one implementation, the user interface instructions 129 may also provide the user with a request for confirmation to complete checkout of a product or service purchase. Examples of the voice commerce interface are illustrated with respect to FIGS. 5, 6A, and 5B.

In an implementation, user interface instructions 129 may generate a profile information management interface. The profile management interface may allow the user to set profile information related to the purchase of products or services. For example, the profile management interface enables the user to set default payment and shipping information. The profile management interface may also enable the user to set preferred sellers, brands, style, size, or other parameters related to products or services as well as limits of certain categories of products or services and particular price ranges.

Examples of System Architectures and Configurations

Different system architectures may be used. For example, all or a portion of voice commerce application 120 may be executed on a user device. In other words, computing device 110 as illustrated may include a user device operated by the user. In implementations where all or a portion of voice commerce application 120 is executed on the user device, the user device may search remote information sources, prepare or complete checkout of product or service purchases, generate the interface, and/or perform other functions/operations of voice commerce application 120.

All or a portion of voice commerce application 120 may be executed on a server device. In other words, computing device 110 as illustrated may include a server device that obtains a user request from a user device operated by the user. In implementations where all or a portion of voice commerce application 120 is executed on the server device, the server device may search remote information sources 140, obtain sets of results, prepare or complete checkout of product or service purchases, provide the interface to a user device, and/or perform other functions/operations of voice commerce application 120.

Although illustrated in FIG. 1 as a single component, computer system 104 may include a plurality of individual components (e.g., computer devices) each programmed with at least some of the functions described herein. In this manner, some components of computer system 104 may perform some functions while other components may perform other functions, as would be appreciated. The one or more processors 112 may each include one or more physical processors that are programmed by computer program instructions. The various instructions described herein are exemplary only. Other configurations and numbers of instructions may be used, so long as the processor(s) 112 are programmed to perform the functions described herein.

It should be appreciated that although the various instructions are illustrated in FIG. 1 as being co-located within a single computing device 110, one or more instructions may be executed remotely from the other instructions. For

US 11,087,385 B2

17

example, some computing devices **110** of computer system **104** may be programmed by some instructions while other computing devices **110** may be programmed by other instructions, as would be appreciated. Furthermore, the various instructions described herein are exemplary only. Other configurations and numbers of instructions may be used, so long as processor(s) **112** are programmed to perform the functions described herein.

The description of the functionality provided by the different instructions described herein is for illustrative purposes, and is not intended to be limiting, as any of instructions may provide more or less functionality than is described. For example, one or more of the instructions may be eliminated, and some or all of its functionality may be provided by other ones of the instructions. As another example, processor(s) **112** may be programmed by one or more additional instructions that may perform some or all of the functionality attributed herein to one of the instructions.

The various instructions described herein may be stored in a storage device **114**, which may comprise random access memory (RAM), read only memory (ROM), and/or other memory. The storage device may store the computer program instructions (e.g., the aforementioned instructions) to be executed by processor(s) **112** as well as data that may be manipulated by processor(s) **112**. The storage device may comprise floppy disks, hard disks, optical disks, tapes, or other storage media for storing computer-executable instructions and/or data.

The various components illustrated in FIG. **1** may be coupled to at least one other component via a network **102**, which may include any one or more of, for instance, the Internet, an intranet, a PAN (Personal Area Network), a LAN (Local Area Network), a WAN (Wide Area Network), a SAN (Storage Area Network), a MAN (Metropolitan Area Network), a wireless network, a cellular communications network, a Public Switched Telephone Network, and/or other network. In FIG. **1** and other drawing Figures, different numbers of entities than depicted may be used. Furthermore, according to various implementations, the components described herein may be implemented in hardware and/or software that configure hardware.

User device(s) may include a device that can interact with computer system **104** through network **102**. Such user device(s) may include, without limitation, a tablet computing device, a smartphone, a laptop computing device, a desktop computing device, a network-enabled appliance such as a "Smart" television, a vehicle computing device, and/or other device that may interact with computer system **104**.

The various databases **132** described herein may be, include, or interface to, for example, an Oracle™ relational database sold commercially by Oracle Corporation. Other databases, such as Informix™, DB2 (Database 2) or other data storage, including file-based (e.g., comma or tab separated files), or query formats, platforms, or resources such as OLAP (On Line Analytical Processing), SQL (Structured Query Language), a SAN (storage area network), Microsoft Access™, MySQL, PostgreSQL, HSpace, Apache Cassandra, MongoDB, Apache CouchDB™, or others may also be used, incorporated, or accessed. The database may comprise one or more such databases that reside in one or more physical devices and in one or more physical locations. The database may store a plurality of types of data and/or files and associated data or file descriptions, administrative information, or any other data. The database(s) **132** may be stored in storage device **114** and/or other storage that is accessible to computer system **104**.

18

Example Flow Diagrams

The following flow diagrams describe operations that may be accomplished using some or all of the system components described in detail above and, in some implementations, various operations may be performed in different sequences and various operations may be omitted. Additional operations may be performed along with some or all of the operations shown in the depicted flow diagrams. One or more operations may be performed simultaneously. Accordingly, the operations as illustrated (and described in greater detail below) are exemplary by nature and, as such, should not be viewed as limiting.

FIG. **3** depicts a data flow diagram for a process **300** for providing voice commerce, according to an implementation of the invention. The various processing operations and/or data flows depicted in FIG. **3** (and in the other drawing figures) are described in greater detail herein. The described operations may be accomplished using some or all of the system components described in detail above and, in some implementations, various operations may be performed in different sequences and various operations may be omitted. Additional operations may be performed along with some or all of the operations shown in the depicted flow diagrams. One or more operations may be performed simultaneously. Accordingly, the operations as illustrated (and described in greater detail below) are exemplary by nature and, as such, should not be viewed as limiting.

In an operation **302**, a natural language utterance of a user may be received. As an example, upon receipt, the natural language utterance may be processed by a speech recognition engine to recognize one or more words of the natural language utterance. The recognized words may then be processed, along with context information associated with the user, by a natural language processing engine to determine one or more products or services and/or other information (e.g., payment information, shipping information, seller information, associated cost information, etc.) that can be used to prepare and/or complete checkout of product or service purchases.

In an operation **304**, a product or service to be purchased on behalf of the user may be determined based on the natural language utterance. As an example, upon recognition of one or more words associated with the natural language utterance that are related to at least one product or service, the product/service-related words may be utilized to search one or more databases associated with one or more third party retailers and/or service providers, and obtain results indicating one or more products or services. The results may then be processed, along with context information associated with the user, to select the most relevant product or service as the product or service that is to be purchased on behalf of the user.

As illustrated by FIG. **3**, in one implementation, the product or service that is to be purchased may be determined without any further user input after the receipt of the natural language utterance (from which the determination of the product or service to be purchased is based). In another implementation, the product or service that is to be purchased may be determined without further user input identifying the product or service after the receipt of the natural language utterance.

In an operation **306**, payment information that is to be used to pay for the product or service may be retrieved. As an example, default payment information associated with the user (e.g., for paying for product or services on behalf of the user) may be stored as profile information associated with the user. Such default payment may, for example, be

US 11,087,385 B2

19

retrieved from one or more databases storing the user's profile information, and utilized to pay for the product or service.

In an operation **308**, shipping information that is to be used to deliver the product or service may be retrieved. As an example, default shipping information associated with the user (e.g., for delivering products or services on behalf of the user) may be stored as profile information associated with the user. Such default shipping information may, for instance, be retrieved from one or more databases storing the user's profile information, and be utilized to deliver the product or service.

In an operation **310**, a purchase transaction for the product or service may be completed based on the payment information and the shipping information.

As illustrated by FIG. 3, in one implementation, the purchase transaction may be completed without any further user input after the receipt of the natural language utterance (from which the determination of the product or service to be purchased is based). In another implementation, the product or service that is to be purchased may be determined without further user input identifying the product or service, identifying the payment information, or identifying the shipping information after the receipt of the natural language utterance.

FIG. 4 depicts a data flow diagram for another process **400** for providing voice commerce, according to an implementation of the invention. The various processing operations and/or data flows depicted in FIG. 4 (and in the other drawing figures) are described in greater detail herein. The described operations may be accomplished using some or all of the system components described in detail above and, in some implementations, various operations may be performed in different sequences and various operations may be omitted. Additional operations may be performed along with some or all of the operations shown in the depicted flow diagrams. One or more operations may be performed simultaneously. Accordingly, the operations as illustrated (and described in greater detail below) are exemplary by nature and, as such, should not be viewed as limiting.

As illustrated by FIG. 3, a natural language utterance of a user may be received, and a product or service that is to be purchased on behalf of the user may be determined based on the natural language utterance without further user input after the receipt of the natural language utterance. Payment information that is to be used to pay for the product or service, and shipping information that is to be used to deliver the product or service, may also be retrieved.

Referring back to FIG. 4, in an operation **402**, a seller from which the product or service is to be purchased may be selected. As an example, the user may have specified a set of sellers from which the user prefers to purchase products or services. As such, the seller may be selected from the predetermined set of sellers specified by the user. As another example, an administrator of the system **100** (or of a subsystem of the system **100**) may have specified a set of sellers that have priority over other sellers in having their products or services offered to users. Thus, the seller may be selected from the predetermined set of sellers specified by the administrator. As yet another example, a database of sellers may be queried to determine which seller has the product or service available for purchase. The sellers that have the product or service available for purchase may then be compared against one another to determine, for example, which seller has the lower price (or overall lower total purchase costs) for the product or service, which sellers has

20

the highest customer rating, or other criteria in selecting the seller from which the product or service is to be purchased.

In an operation **404**, seller information associated with the selected seller may be retrieved. As an example, seller information may comprise a name of the seller, contact information associated with the seller, a price at which the seller will sell the product or service to the user, individual or overall costs involved with purchase of the product or service from the seller, a refund policy of the seller for the product or service, or other information.

In an operation **406**, a request for user confirmation (to purchase the product or service from the seller and to use the payment information and the shipping information for a purchase transaction for the product or service) may be provided without further user input after the receipt of the utterance.

In an operation **408**, it may be determined that the user has confirmed purchasing of the product or service from the seller and use of the payment information and the shipping information for the purchase transaction. As such, in an operation **410**, the purchase transaction may be completed based on the seller information, the payment information, and the shipping information without further user input after the receipt of the utterance other than a user confirmation with respect to the confirmation request.

Example Screenshots

FIG. 5 illustrates a screenshot **502** of a user interface which facilitates voice commerce, according to an implementation of the invention. The described screenshots may be accomplished using some or all of the system components described in detail above. As shown in screenshot **502** of FIG. 5, a voice commerce application may enable the user to complete checkout of product or service purchases via a single utterance. As depicted in screenshot **502**, the user may say "Send flowers to Betsy." In response to receiving the utterance from the user, the voice commerce application may determine that flowers are to be purchased, and select a third party retailer from which to purchase the flowers. As an example, the user may have purchased a particular kind of flower (e.g., red roses) for Betsy numerous times in the past. As such, the particular kind of flower may be determined as the particular product that is to be purchased on behalf of the user for Betsy. The selection of the product may, for example, be performed without further user input identifying the product type (e.g., flowers) after the receipt of the utterance, without any further user input after the receipt of the utterance, etc.

The voice commerce application may also determine the payment information to be used to pay for the flowers. Because the utterance indicates that the flowers are to be sent to Betsy, the voice commerce application may determine Betsy's address from the user's address book or contact list and provide Betsy's shipping information to the retailer. The voice commerce application may then, for example, complete a purchase transaction using seller information associated with the third party retailer, the payment information, the shipping information, or other information without further user input from the user. Upon completing the purchase transaction for the flowers, the voice commerce application may indicate that checkout for the flowers has been completed. In an implementation, the product to be purchased, a seller from which the product is to be purchased, payment information, shipping information, or other information may be displayed to the user.

FIGS. 6A and 6B illustrate screenshots **602** and **604** of another user interface which facilitates voice commerce, according to an implementation of the invention. The

US 11,087,385 B2

21

described screenshots may be accomplished using some or all of the system components described in detail above. As shown in screenshot **602** of FIG. **6A**, a voice commerce application may solicit a user confirmation from a user with respect to a product purchase before completing checkout of the product purchase. As depicted in screenshot **602**, the user may say "Send flowers to Betsy." In response to receiving the single utterance from the user, the voice commerce application may determine that flowers are to be purchased, and select a third party retailer from which to purchase the flowers. The voice commerce application may also determine the payment information to be used to pay for the flowers. Because the utterance indicates that the flowers are to be sent to Betsy, the voice commerce application may determine Betsy's address from the user's address book or contact list and provide Betsy's shipping information to the retailer. The voice commerce application may then present a request for user confirmation of the purchase to the user (e.g., without further user input after the receipt of the utterance). As shown in screenshot **604**, the user may confirm purchase of the flowers by saying "Yes." Upon confirmation of the purchase, the voice commerce application may indicate that checkout for the flowers has been completed.

FIG. **7A** illustrates a screenshot **702** of a user interface which facilitates management of user profile information, according to an implementation of the invention. The described screenshots may be accomplished using some or all of the system components described in detail above. As shown in screenshot **702**, the user profile information interface includes the user name, default payment information, and default shipping information. The user may edit any data fields associated with the user name, default payment information, and default shipping information.

FIG. **7B** illustrates a screenshot **704** of another user interface which facilitates management of user profile information, according to an implementation of the invention. The described screenshots may be accomplished using some or all of the system components described in detail above. As shown in screenshot **704**, the user profile information interface may enable the user to pre-set preferences of different products and/or services to be purchased. The profile information interface includes a product/service category that may be selected by the user. For each product/service category, the user may select preferred sellers, preferred brands, preferred styles, preferred sizes, and other parameters related to the products/services to be selected. The user may also limit certain categories of products or services to a particular price range. The user may edit any data field associated with product/service category to include preferred retail stores/service providers, preferred brands, preferred styles, preferred sizes, and price ranges.

Other implementations, uses and advantages of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. The specification should be considered exemplary only, and the scope of the invention is accordingly intended to be limited only by the following claims.

What is claimed is:

1. A method for providing voice commerce, the method being implemented on a computer system having one or more physical processors programmed with computer program instructions which, when executed, perform the method, the method comprising:

receiving, by the computer system, a single first user input comprising a natural language utterance;

22

providing, by the computer system, the natural language utterance as an input to a speech recognition engine; obtaining, by the computer system, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine;

searching, by the computer system, one or more databases of products or services based on the one or more words or phrases;

selecting, by the computer system, without further user input other than the single first user input, a product or service from the database to be purchased based on the search;

receiving, by the computer system, a second user input indicating confirmation by a user to complete a purchase transaction of the selected product or service; and completing, by the computer system, without further user input after the receipt of the second user input, a purchase transaction of the selected product or service.

2. The method of claim 1, wherein selecting the product or service further comprises:

determining, by the computer system, a context based at least on the one or more words or phrases, wherein the product or service is selected based at least on the determined context.

3. The method of claim 1, wherein completing the purchase transaction for the selected product or service comprises:

obtaining, by the computer system, payment information with which to pay for the selected product or service, wherein the purchase transaction is completed based on the payment information without receiving confirmation of the payment information by the user.

4. The method of claim 1, wherein completing the purchase transaction for the selected product or service comprises:

obtaining, by the computer system, shipping information with which to deliver the selected product or service, wherein the shipping information specifies a name or address of a recipient to which the selected product or service is to be delivered after the selected product or service is purchased, and wherein the purchase transaction is completed based on the shipping information without receiving confirmation of the shipping information by the user.

5. The method of claim 1, the method further comprising: obtaining, by the computer system, seller information describing one or more products or services available from one or more sellers via one or more remote information sources; and

storing, by the computer system, the seller information in the one or more databases.

6. The method of claim 5, wherein the one or more remote information sources comprise at least a third party search engine, a third party retailer, and/or a third party service provider.

7. The method of claim 5, wherein the method further comprises:

selecting, by the computer system, a seller from which to purchase the selected product or service.

8. The method of claim 7, wherein the seller information comprises a price at which the one or more sellers will sell the one or more products or services, wherein selecting the seller comprises selecting the seller based on the price at which the seller will sell the selected product or service.

9. The method of claim 7, the method further comprising: obtaining, by the computer system, user profile information associated with the user, wherein the user profile

US 11,087,385 B2

23

information indicates a predetermined set of sellers associated with the user, wherein selecting the seller comprises selecting the seller from the predetermined set of sellers indicated by the user profile information.

10. The method of claim 7, the method further comprising:

obtaining, by the computer system, a predetermined set of sellers specified by an administrator of the system that is different than the user, wherein selecting the seller comprises selecting the seller from the predetermined set of administrator-specified sellers.

11. The method of claim 1, wherein completing the purchase transaction of the selected product or service comprises:

obtaining, by the computer system, payment information with which to pay for the selected product or service; and

obtaining, by the computer system, shipping information with which to deliver the selected product or service, wherein the shipping information specifies a name or address of a recipient to which the selected product or service is to be delivered after the selected product or service is purchased, and wherein the purchase transaction is completed based on the payment information and the shipping information.

12. The method of claim 11, the method further comprising:

identifying, by the computer system, an intended recipient of the identified product or service based on the single first user input and/or the second user input, and wherein obtaining the shipping information comprises: obtaining, by the computer system, an address of the intended recipient.

13. The method of claim 12, wherein obtaining the address of the intended recipient comprises:

accessing, by the computer system, an address book of the user, wherein the address book comprises an identification of the intended recipient and the address of the intended recipient.

14. The method of claim 11, the method further comprising:

completing, by the computer system, the purchase transaction without receiving confirmation of the payment information or the shipping information by the user.

15. The method of claim 1, the method further comprising:

providing, by the computer system, without further user input after the receipt of other than the single first user input, a request for user confirmation to complete the purchase transaction for the selected product or service, wherein the second user input is received responsive to the request;

determining, by the computer system, that the user has confirmed the purchase transaction based on the second user input, wherein the purchase transaction of the selected product or service is completed based on the determination.

16. A system for providing voice commerce, the system comprising:

one or more physical processors programmed with computer program instructions which, when executed, cause the one or more physical processors to:

receive a single first user input comprising a natural language utterance;

provide the natural language utterance as an input to a speech recognition engine;

24

obtain one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine;

search one or more databases of products or services based on the one or more words or phrases;

select, without further user input other than the single first user input, a product or service from the database to be purchased based on the search;

receive a second user input indicating confirmation by a user to complete a purchase transaction of the selected product or service; and

complete, without further user input after the receipt of the second user input, a purchase transaction of the selected product or service.

17. The system of claim 16, wherein to select the product or service, the one or more physical processors are further caused to:

determine a context based at least on the one or more words or phrases, wherein the product or service is selected based at least on the determined context.

18. The system of claim 16, wherein to complete the purchase transaction for the selected product or service, the one or more physical processors are further caused to:

obtain, without further user input after the receipt of the user input, payment information with which to pay for the selected product or service, wherein the purchase transaction is completed based on the payment information without receiving confirmation of the payment information by the user.

19. The system of claim 16, wherein to complete the purchase transaction for the selected product or service, the one or more physical processors are further caused to:

obtain, shipping information with which to deliver the selected product or service, wherein the shipping information specifies a name or address of a recipient to which the selected product or service is to be delivered after the selected product or service is purchased, and wherein the purchase transaction is completed based on the shipping information without receiving confirmation of the shipping information by the user.

20. The system of claim 16, wherein the one or more physical processors are further caused to:

obtain seller information describing one or more products or services available from one or more sellers via one or more remote information sources; and store the seller information in the one or more databases.

21. The system of claim 20, wherein the one or more remote information sources comprise at least a third party search engine, a third party retailer, and/or a third party service provider.

22. The system of claim 20, wherein the one or more physical processors are further caused to:

select a seller from which to purchase the selected product or service.

23. The system of claim 22, wherein the seller information comprises a price at which the one or more sellers will sell the one or more products or services, wherein to select the seller, the one or more physical processors are further caused to:

select the seller based on the price at which the seller will sell the selected product or service.

24. The system of claim 22, wherein the one or more physical processors are further caused to:

obtain user profile information associated with the user, wherein the user profile information indicates a predetermined set of sellers associated with the user, wherein

US 11,087,385 B2

25

selecting the seller comprises selecting the seller from the predetermined set of sellers indicated by the user profile information.

25. The system of claim 22, wherein the one or more physical processors are further caused to:

obtain a predetermined set of sellers specified by an administrator of the system that is different than the user, wherein selecting the seller comprises selecting the seller from the predetermined set of administrator-specified sellers.

26. The system of claim 16, wherein to complete the purchase transaction of the selected product or service, the one or more physical processors are further caused to:

obtain payment information with which to pay for the selected product or service; and

obtain shipping information with which to deliver the selected product or service, wherein the shipping information specifies a name or address of a recipient to which the selected product or service is to be delivered after the selected product or service is purchased, and wherein the purchase transaction is completed based on the payment information and the shipping information.

27. The system of claim 26, wherein the one or more physical processors are further caused to:

identify an intended recipient of the identified product or service based on the single first user input and/or the second user input, and wherein to obtain the shipping information, the one or more physical processors are further caused to:

obtain an address of the intended recipient.

28. The system of claim 27, wherein to obtain the address of the intended recipient, the one or more physical processors are further caused to:

access an address book of the user, wherein the address book comprises an identification of the intended recipient and the address of the intended recipient.

29. The system of claim 26, wherein the one or more physical processors are further caused to:

complete the purchase transaction without receiving confirmation of the payment information or the shipping information by the user.

30. The system of claim 16, wherein the one or more physical processors are further caused to:

provide, without further user input other than the single first user input, a request for user confirmation to complete the purchase transaction for the selected product or service, wherein the second user input is received responsive to the request;

determine that the user has confirmed the purchase transaction based on the second user input, wherein the purchase transaction of the selected product or service is completed based on the determination.

31. A method for providing voice commerce, the method being implemented on a computer system having one or more physical processors programmed with computer program instructions which, when executed, perform the method, the method comprising:

receiving, by the computer system, a single first user input comprising a natural language utterance;

recognizing, by the computer system, one or more words or phrases from the natural language utterance;

searching, by the computer system, one or more databases of products or services based on the one or more recognized words or phrases from the single first user input, and without using further user input other than the single first user input;

26

causing, by the computer system, a set of search results to be presented to a user based on the search, the search results indicating one or more products or services from the database available for purchase;

receiving, by the computer system, a second user input comprising a selection from the set of search results, the selection identifying one or more products or services from the database to be purchased on behalf of the user based on the second user input;

obtaining, by the computer system, user profile information associated with the user;

identifying, by the computer system, payment information and shipping information based on the user profile information; and

completing, by the computer system, without further user input after identifying the payment information and the shipping information, a purchase transaction of the identified one or more products or services.

32. The method of claim 31, wherein recognizing the one or more words or phrases from the natural language utterance comprises:

providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and

obtaining, by the computer system, the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.

33. The method of claim 31, the method further comprising:

obtaining, by the computer system, seller information describing one or more products or services available from one or more sellers via one or more remote information sources; and

storing, by the computer system, the seller information in the one or more databases.

34. The method of claim 31, wherein completing the purchase transaction without further user input after identifying the payment information and the shipping information comprises:

completing, by the computer system, the purchase transaction without receiving confirmation of the payment information or the shipping information by the user.

35. A system for providing voice commerce, the system comprising:

one or more physical processors programmed with computer program instructions which, when executed, cause the one or more physical processors to:

receive a single first user input comprising a natural language utterance;

recognize one or more words or phrases from the natural language utterance;

search one or more databases of products or services based on the one or more recognized words or phrases from the single first user input, and without using further user input other than the single first user input;

cause a set of search results to be presented to a user based on the search, the search results indicating one or more products or services from the database available for purchase;

receive a second user input comprising a selection from the set of search results, the selection identifying one or more products or services from the database to be

27

purchased on behalf of the user based on the second user input;
obtain user profile information associated with the user;
identify payment information and shipping information based on the user profile information; and
complete, without further user input after identifying the payment information and the shipping information, a purchase transaction of the identified one or more products or services.
36. The system of claim 35, wherein to recognize the one or more words or phrases from the natural language utterance, the one or more physical processors are further caused to:
provide the natural language utterance as an input to a speech recognition engine; and
obtain the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.
37. The system of claim 35, wherein the one or more physical processors are further caused to:
obtain seller information describing one or more products or services available from one or more sellers via one or more remote information sources; and
store the seller information in the one or more databases.

28

38. The system of claim 35, wherein to complete the purchase transaction without further user input after identifying the payment information and the shipping information, the one or more physical processors are further caused to:
complete the purchase transaction without receiving confirmation of the payment information or the shipping information by the user.
39. The method of claim 1, the method further comprising:
presenting a prompt that identifies the selected product or service, the cost associated with the purchase of the selected product or service, payment information to pay the associated cost, and shipping information specifying where the selected product or service is to be delivered; and
soliciting approval of the identified information as the second user input.
40. The system of claim 16, wherein the one or more physical processors to are further caused to:
present a prompt that identifies the selected product or service, the cost associated with the purchase of the selected product or service, payment information to pay the associated cost, and shipping information specifying where the selected product or service is to be delivered; and
solicit approval of the identified information as the second user input.

* * * * *

EXHIBIT D



US011080758B2

(12) **United States Patent**
Freeman et al.

(10) **Patent No.:** **US 11,080,758 B2**

(45) **Date of Patent:** **Aug. 3, 2021**

(54) **SYSTEM AND METHOD FOR DELIVERING TARGETED ADVERTISEMENTS AND/OR PROVIDING NATURAL LANGUAGE PROCESSING BASED ON ADVERTISEMENTS**

(71) Applicant: **VB Assets, LLC**, Bellevue, WA (US)

(72) Inventors: **Tom Freeman**, Mercer Island, WA (US); **Mike Kennewick**, Bellevue, WA (US)

(73) Assignee: **VB Assets, LLC**, Bellevue, WA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **16/194,944**

(22) Filed: **Nov. 19, 2018**

(65) **Prior Publication Data**

US 2019/0087866 A1 Mar. 21, 2019

Related U.S. Application Data

(63) Continuation of application No. 15/223,870, filed on Jul. 29, 2016, now Pat. No. 10,134,060, which is a (Continued)

(51) **Int. Cl.**
G06Q 30/02 (2012.01)
G10L 15/26 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC **G06Q 30/0269** (2013.01); **G06F 40/263** (2020.01); **G06Q 30/0241** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC G06Q 30/0251; G06F 17/27; G10L 15/22
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,430,669 A 2/1984 Cheung
4,821,027 A 4/1989 Mallory
(Continued)

FOREIGN PATENT DOCUMENTS

CN 1433554 A 7/2003
CN 1860496 A 11/2006
(Continued)

OTHER PUBLICATIONS

"Statement in Accordance with the Notice from the European Patent Office" dated Oct. 1, 2007 Concerning Business Methods (OJ EPO Nov. 2007, 592-593), XP002456252.

(Continued)

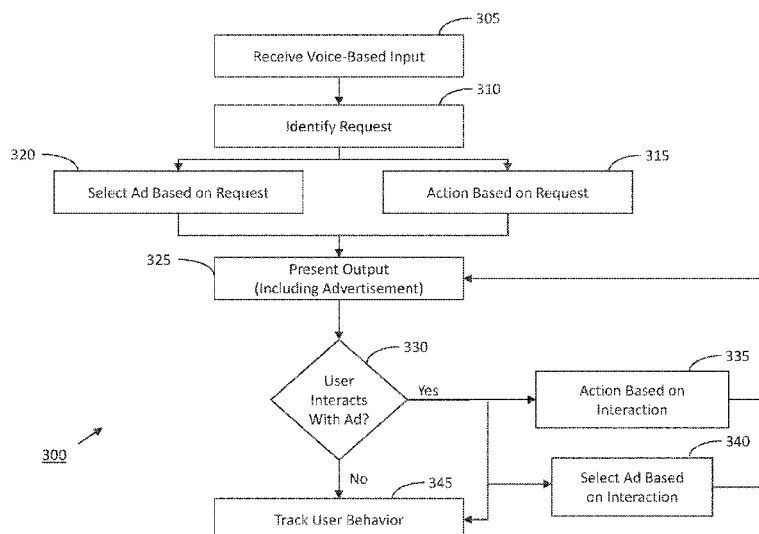
Primary Examiner — Brian L. Albertalli

(74) *Attorney, Agent, or Firm* — Sheppard Mullin Richter & Hampton LLP

(57) **ABSTRACT**

The system and method described herein may use various natural language models to deliver targeted advertisements and/or provide natural language processing based on advertisements. In one implementation, an advertisement associated with a product or service may be provided for presentation to a user. A natural language utterance of the user may be received. The natural language utterance may be interpreted based on the advertisement and, responsive to the existence of a pronoun in the natural language utterance, a determination of whether the pronoun refers to one or more of the product or service or a provider of the product or service may be effectuated.

44 Claims, 3 Drawing Sheets



US 11,080,758 B2

Page 2

Related U.S. Application Data

continuation of application No. 14/836,606, filed on Aug. 26, 2015, now Pat. No. 9,406,078, which is a continuation of application No. 14/537,598, filed on Nov. 10, 2014, now Pat. No. 9,269,097, which is a continuation of application No. 14/016,757, filed on Sep. 3, 2013, now Pat. No. 8,886,536, which is a continuation of application No. 13/371,870, filed on Feb. 13, 2012, now Pat. No. 8,527,274, which is a continuation of application No. 12/847,564, filed on Jul. 30, 2010, now Pat. No. 8,145,489, which is a continuation of application No. 11/671,526, filed on Feb. 6, 2007, now Pat. No. 7,818,176.

(51) Int. Cl.

G06F 40/263 (2020.01)

G10L 15/18 (2013.01)

(52) U.S. Cl.

CPC G06Q 30/0242 (2013.01); G06Q 30/0251 (2013.01); G10L 15/18 (2013.01); G10L 15/26 (2013.01)

(56) References Cited

U.S. PATENT DOCUMENTS

4,829,423	A	5/1989	Tennant	5,761,631	A	6/1998	Nasukawa
4,887,212	A	12/1989	Zamora	5,774,841	A	6/1998	Salazar
4,910,784	A	3/1990	Doddington	5,774,859	A	6/1998	Houser
5,027,406	A	6/1991	Roberts	5,794,050	A	8/1998	Dahlgren
5,155,743	A	10/1992	Jacobs	5,794,196	A	8/1998	Yegnaranayanan
5,164,904	A	11/1992	Sumner	5,797,112	A	8/1998	Komatsu
5,208,748	A	5/1993	Flores	5,799,276	A	8/1998	Komissarchik
5,265,065	A	11/1993	Turtle	5,802,510	A	9/1998	Jones
5,274,560	A	12/1993	LaRue	5,829,000	A	10/1998	Huang
5,331,554	A	7/1994	Graham	5,832,221	A	11/1998	Jones
5,357,596	A	10/1994	Takebayashi	5,839,107	A	11/1998	Gupta
5,369,575	A	11/1994	Lamberti	5,848,396	A	12/1998	Gerace
5,377,350	A	12/1994	Skinner	5,855,000	A	12/1998	Waibel
5,386,556	A	1/1995	Hedin	5,860,059	A	1/1999	Aust
5,424,947	A	6/1995	Nagao	5,867,817	A	2/1999	Catallo
5,471,318	A	11/1995	Ahuja	5,878,385	A	3/1999	Bralich
5,475,733	A	12/1995	Eisdorfer	5,878,386	A	3/1999	Coughlin
5,479,563	A	12/1995	Yamaguchi	5,892,813	A	4/1999	Morin
5,488,652	A	1/1996	Bielby	5,892,900	A	4/1999	Ginter
5,499,289	A	3/1996	Bruno	5,895,464	A	4/1999	Bhandari
5,500,920	A	3/1996	Kupiec	5,895,466	A	4/1999	Goldberg
5,517,560	A	5/1996	Greenspan	5,897,613	A	4/1999	Chan
5,533,108	A	7/1996	Harris	5,898,760	A	4/1999	Smets
5,537,436	A	7/1996	Bottoms	5,899,991	A	5/1999	Karch
5,539,744	A	7/1996	Chu	5,902,347	A	5/1999	Backman
5,557,667	A	9/1996	Bruno	5,911,120	A	6/1999	Jarett
5,559,864	A	9/1996	Kennedy, Jr.	5,918,222	A	6/1999	Fukui
5,563,937	A	10/1996	Bruno	5,926,784	A	7/1999	Richardson
5,577,165	A	11/1996	Takebayashi	5,933,822	A	8/1999	Braden-Harder
5,590,039	A	12/1996	Ikeda	5,950,167	A	9/1999	Yaker
5,608,635	A	3/1997	Tamai	5,953,393	A	9/1999	Culbreth
5,615,296	A	3/1997	Stanford	5,960,384	A	9/1999	Brash
5,617,407	A	4/1997	Bareis	5,960,397	A	9/1999	Rahim
5,633,922	A	5/1997	August	5,960,399	A	9/1999	Barclay
5,634,086	A	5/1997	Rtischev	5,960,447	A	9/1999	Holt
5,652,570	A	7/1997	Lepkofker	5,963,894	A	10/1999	Richardson
5,675,629	A	10/1997	Raffel	5,963,940	A	10/1999	Liddy
5,696,965	A	12/1997	Dedrick	5,982,906	A	11/1999	Ono
5,708,422	A	1/1998	Blonder	5,983,190	A	11/1999	Trower, II
5,721,938	A	2/1998	Stuckey	5,987,404	A	11/1999	DellaPietra
5,722,084	A	2/1998	Chakrin	5,991,721	A	11/1999	Asano
5,729,659	A	3/1998	Potter	5,995,119	A	11/1999	Cosatto
5,740,256	A	4/1998	CastelloDaCosta	5,995,928	A	11/1999	Nguyen
5,742,763	A	4/1998	Jones	5,995,943	A	11/1999	Bull
5,748,841	A	5/1998	Morin	6,009,382	A	12/1999	Martino
5,748,974	A	5/1998	Johnson	6,014,559	A	1/2000	Amin
5,752,052	A	5/1998	Richardson	6,018,708	A	1/2000	Dahan
5,754,784	A	5/1998	Garland	6,021,384	A	2/2000	Gorin
				6,028,514	A	2/2000	Lemelson
				6,035,267	A	3/2000	Watanabe
				6,044,347	A	3/2000	Abella
				6,049,602	A	4/2000	Foladare
				6,049,607	A	4/2000	Marash
				6,058,187	A	5/2000	Chen
				6,067,513	A	5/2000	Ishimitsu
				6,073,098	A	6/2000	Buchsbaum
				6,076,057	A	6/2000	Narayanan
				6,076,059	A	6/2000	Glickman
				6,078,886	A	6/2000	Dragosh
				6,081,774	A	6/2000	deHita
				6,081,779	A	6/2000	Besling
				6,085,186	A	7/2000	Christianson
				6,101,241	A	8/2000	Boyce
				6,108,631	A	8/2000	Ruhl
				6,119,087	A	9/2000	Kuhn
				6,119,101	A	9/2000	Peckover
				6,122,613	A	9/2000	Baker
				6,134,235	A	10/2000	Goldman
				6,144,667	A	11/2000	Doshi
				6,144,938	A	11/2000	Surace
				6,154,526	A	11/2000	Dahlke
				6,160,883	A	12/2000	Jackson
				6,167,377	A	12/2000	Gillick
				6,173,250	B1	1/2001	Jong
				6,173,266	B1	1/2001	Marx
				6,173,279	B1	1/2001	Levin
				6,175,858	B1	1/2001	Bulfer
				6,185,535	B1	2/2001	Hedin
				6,188,982	B1	2/2001	Chiang

US 11,080,758 B2

Page 3

(56)

References Cited

U.S. PATENT DOCUMENTS

6,192,110 B1	2/2001	Abella	6,501,833 B2	12/2002	Phillips
6,192,338 B1	2/2001	Haszto	6,501,834 B1	12/2002	Milewski
6,195,634 B1	2/2001	Dudemaine	6,505,155 B1	1/2003	Vanbuskirk
6,195,651 B1	2/2001	Handel	6,510,417 B1	1/2003	Woods
6,199,043 B1	3/2001	Happ	6,513,006 B2	1/2003	Howard
6,208,964 B1	3/2001	Sabourin	6,522,746 B1	2/2003	Marchok
6,208,972 B1	3/2001	Grant	6,523,061 B1	2/2003	Halverson
6,219,346 B1	4/2001	Maxemchuk	6,526,335 B1	2/2003	Treyz
6,219,643 B1	4/2001	Cohen	6,532,444 B1	3/2003	Weber
6,219,645 B1	4/2001	Byers	6,539,348 B1	3/2003	Bond
6,226,612 B1	5/2001	Srenger	6,549,629 B2	4/2003	Finn
6,233,556 B1	5/2001	Teunen	6,553,372 B1	4/2003	Brassell
6,233,559 B1	5/2001	Balakrishnan	6,556,970 B1	4/2003	Sasaki
6,233,561 B1	5/2001	Junqua	6,556,973 B1	4/2003	Lewin
6,236,968 B1	5/2001	Kanevsky	6,560,576 B1	5/2003	Cohen
6,243,679 B1	6/2001	Mohri	6,560,590 B1	5/2003	Shwe
6,246,981 B1	6/2001	Papineni	6,567,778 B1	5/2003	ChaoChang
6,246,990 B1	6/2001	Happ	6,567,797 B1	5/2003	Schuetze
6,266,636 B1	7/2001	Kosaka	6,567,805 B1	5/2003	Johnson
6,269,336 B1	7/2001	Ladd	6,570,555 B1	5/2003	Prevost
6,272,455 B1	8/2001	Hoshen	6,570,964 B1	5/2003	Murveit
6,272,461 B1	8/2001	Meredith	6,571,279 B1	5/2003	Herz
6,275,231 B1	8/2001	Obradovich	6,574,597 B1	6/2003	Mohri
6,278,377 B1	8/2001	DeLine	6,574,624 B1	6/2003	Johnson
6,278,968 B1	8/2001	Franz	6,578,022 B1	6/2003	Foulger
6,286,002 B1	9/2001	Axaopoulos	6,581,103 B1	6/2003	Dengler
6,288,319 B1	9/2001	Catona	6,584,439 B1	6/2003	Geilhufe
6,292,767 B1	9/2001	Jackson	6,587,858 B1	7/2003	Strazza
6,301,560 B1	10/2001	Masters	6,591,185 B1	7/2003	Polidi
6,308,151 B1	10/2001	Smith	6,591,239 B1	7/2003	McCall
6,311,159 B1	10/2001	VanTichelen	6,594,257 B1	7/2003	Doshi
6,314,402 B1	11/2001	Monaco	6,594,367 B1	7/2003	Marash
6,321,196 B1	11/2001	Franceschi	6,598,018 B1	7/2003	Junqua
6,356,869 B1	3/2002	Chapados	6,601,026 B2	7/2003	Appelt
6,362,748 B1	3/2002	Huang	6,601,029 B1	7/2003	Pickering
6,366,882 B1	4/2002	Bijl	6,604,075 B1	8/2003	Brown
6,366,886 B1	4/2002	Dragosh	6,604,077 B2	8/2003	Dragosh
6,374,214 B1	4/2002	Friedland	6,606,598 B1	8/2003	Holthouse
6,374,226 B1	4/2002	Hunt	6,611,692 B2	8/2003	Raffel
6,377,913 B1	4/2002	Coffman	6,614,773 B1	9/2003	Maxemchuk
6,377,919 B1	4/2002	Burnett	6,615,172 B1	9/2003	Bennett
6,381,535 B1	4/2002	Durocher	6,622,119 B1	9/2003	Ramaswamy
6,385,596 B1	5/2002	Wiser	6,629,066 B1	9/2003	Jackson
6,385,646 B1	5/2002	Brown	6,631,346 B1	10/2003	Karaorman
6,389,398 B1	5/2002	Lustgarten	6,631,351 B1	10/2003	Ramachandran
6,393,403 B1	5/2002	Majaniemi	6,633,846 B1	10/2003	Bennett
6,393,428 B1	5/2002	Miller	6,636,790 B1	10/2003	Lightner
6,397,181 B1	5/2002	Li	6,643,620 B1	11/2003	Contolini
6,404,878 B1	6/2002	Jackson	6,647,363 B2	11/2003	Claassen
6,405,170 B1	6/2002	Phillips	6,650,747 B1	11/2003	Bala
6,408,272 B1	6/2002	White	6,658,388 B1	12/2003	Kleindienst
6,411,810 B1	6/2002	Maxemchuk	6,678,680 B1	1/2004	Woo
6,411,893 B2	6/2002	Ruhl	6,681,206 B1	1/2004	Gorin
6,415,257 B1	7/2002	Junqua	6,691,151 B1	2/2004	Cheyser
6,418,210 B1	7/2002	Sayko	6,701,294 B1	3/2004	Ball
6,420,975 B1	7/2002	DeLine	6,704,396 B2	3/2004	Parolkar
6,429,813 B2	8/2002	Feigen	6,704,576 B1	3/2004	Brachman
6,430,285 B1	8/2002	Bauer	6,704,708 B1	3/2004	Pickering
6,430,531 B1	8/2002	Polish	6,707,421 B1	3/2004	Drury
6,434,523 B1	8/2002	Monaco	6,708,150 B1	3/2004	Hirayama
6,434,524 B1	8/2002	Weber	6,721,001 B1	4/2004	Berstis
6,434,529 B1	8/2002	Walker	6,721,633 B2	4/2004	Funk
6,442,522 B1	8/2002	Carberry	6,721,706 B1	4/2004	Strubbe
6,446,114 B1	9/2002	Bulfer	6,726,636 B2	4/2004	DerGhazarian
6,453,153 B1	9/2002	Bowker	6,732,088 B1	5/2004	Glance
6,453,292 B2	9/2002	Ramaswamy	6,735,592 B1	5/2004	Neumann
6,456,711 B1	9/2002	Cheung	6,739,556 B1	5/2004	Langston
6,456,974 B1	9/2002	Baker	6,741,931 B1	5/2004	Kohut
6,466,654 B1	10/2002	Cooper	6,742,021 B1	5/2004	Halverson
6,466,899 B1	10/2002	Yano	6,745,161 B1	6/2004	Arnold
6,470,315 B1	10/2002	Netsch	6,751,591 B1	6/2004	Gorin
6,487,494 B2	11/2002	Odinak	6,751,612 B1	6/2004	Schuetze
6,487,495 B1	11/2002	Gale	6,754,485 B1	6/2004	Obradovich
6,498,797 B1	12/2002	Anerousis	6,754,627 B2	6/2004	Woodward
6,499,013 B1	12/2002	Weber	6,754,647 B1	6/2004	Tackett
			6,757,362 B1	6/2004	Cooper
			6,757,544 B2	6/2004	Rangarajan
			6,757,718 B1	6/2004	Halverson
			6,785,651 B1	8/2004	Wang

US 11,080,758 B2

Page 4

(56)

References Cited

U.S. PATENT DOCUMENTS

6,795,808	B1	9/2004	Strubbe	7,137,126	B1	11/2006	Coffman
6,801,604	B2	10/2004	Maes	7,143,037	B1	11/2006	Chestnut
6,801,893	B1	10/2004	Backfried	7,143,039	B1	11/2006	Stifelman
6,804,330	B1	10/2004	Jones	7,146,315	B2	12/2006	Balan
6,810,375	B1	10/2004	Ejerhed	7,146,319	B2	12/2006	Hunt
6,813,341	B1	11/2004	Mahoney	7,149,696	B2	12/2006	Shimizu
6,816,830	B1	11/2004	Kempe	7,165,028	B2	1/2007	Gong
6,823,308	B2	11/2004	Keiller	7,170,993	B2	1/2007	Anderson
6,829,603	B1	12/2004	Chai	7,171,291	B2	1/2007	Obradovich
6,832,230	B1	12/2004	Zilliacus	7,174,300	B2	2/2007	Bush
6,833,848	B1	12/2004	Wolff	7,177,798	B2	2/2007	Hsu
6,839,896	B2	1/2005	Coffman	7,184,957	B2	2/2007	Brookes
6,850,603	B1	2/2005	Eberle	7,190,770	B2	3/2007	Ando
6,856,990	B2	2/2005	Barile	7,197,069	B2	3/2007	Agazzi
6,859,776	B1	2/2005	Cohen	7,197,460	B1	3/2007	Gupta
6,865,481	B2	3/2005	Kawazoe	7,203,644	B2	4/2007	Anderson
6,868,380	B2	3/2005	Kroeker	7,206,418	B2	4/2007	Yang
6,868,385	B1	3/2005	Gerson	7,207,011	B2	4/2007	Mulvey
6,871,179	B1	3/2005	Kist	7,215,941	B2	5/2007	Beckmann
6,873,837	B1	3/2005	Yoshioka	7,228,276	B2	6/2007	Omote
6,877,001	B2	4/2005	Wolf	7,231,343	B1	6/2007	Treadgold
6,877,134	B1	4/2005	Fuller	7,236,923	B1	6/2007	Gupta
6,882,970	B1	4/2005	Garner	7,254,482	B2	8/2007	Kawasaki
6,901,366	B1	5/2005	Kuhn	7,272,212	B2	9/2007	Eberle
6,910,003	B1	6/2005	Arnold	7,277,854	B2	10/2007	Bennett
6,912,498	B2	6/2005	Stevens	7,283,829	B2	10/2007	Christenson
6,915,126	B2	7/2005	Mazzara, Jr.	7,283,951	B2	10/2007	Marchisio
6,928,614	B1	8/2005	Everhart	7,289,606	B2	10/2007	Sibal
6,934,756	B2	8/2005	Maes	7,299,186	B2	11/2007	Kuzunuki
6,937,977	B2	8/2005	Gerson	7,301,093	B2	11/2007	Sater
6,937,982	B2	8/2005	Kitaoka	7,305,381	B1	12/2007	Poppink
6,941,266	B1	9/2005	Gorin	7,321,850	B2	1/2008	Wakita
6,944,594	B2	9/2005	Busayapongchai	7,328,155	B2	2/2008	Endo
6,950,821	B2	9/2005	Faybishenko	7,337,116	B2	2/2008	Charlesworth
6,954,755	B2	10/2005	Reisman	7,340,040	B1	3/2008	Saylor
6,959,276	B2	10/2005	Droppo	7,366,285	B2	4/2008	Parolkar
6,961,700	B2	11/2005	Mitchell	7,366,669	B2	4/2008	Nishitani
6,963,759	B1	11/2005	Gerson	7,376,586	B1	5/2008	Partovi
6,964,023	B2	11/2005	Maes	7,376,645	B2	5/2008	Bernard
6,968,311	B2	11/2005	Knockeart	7,380,250	B2	5/2008	Schechter
6,973,387	B2	12/2005	Masclet	7,386,443	B1	6/2008	Parthasarathy
6,975,983	B1	12/2005	Fortescue	7,398,209	B2	7/2008	Kennewick
6,975,993	B1	12/2005	Keiller	7,406,421	B2	7/2008	Odinak
6,980,092	B2	12/2005	Turnbull	7,415,100	B2	8/2008	Cooper
6,983,055	B2	1/2006	Luo	7,415,414	B2	8/2008	Azara
6,990,513	B2	1/2006	Belfiore	7,421,393	B1	9/2008	DiFabrizio
6,996,531	B2	2/2006	Korall	7,424,431	B2	9/2008	Greene
7,003,463	B1	2/2006	Maes	7,447,635	B1	11/2008	Konopka
7,016,849	B2	3/2006	Arnold	7,451,088	B1	11/2008	Ehlen
7,020,609	B2	3/2006	Thrift	7,454,368	B2	11/2008	Stillman
7,024,364	B2	4/2006	Guerra	7,454,608	B2	11/2008	Gopalakrishnan
7,027,586	B2	4/2006	Bushey	7,461,059	B2	12/2008	Richardson
7,027,974	B1	4/2006	Busch	7,472,020	B2	12/2008	Brulle-Drews
7,027,975	B1	4/2006	Pazandak	7,472,060	B1	12/2008	Gorin
7,031,908	B1	4/2006	Huang	7,472,075	B2	12/2008	Odinak
7,035,415	B2	4/2006	Belt	7,477,909	B2	1/2009	Roth
7,036,128	B1	4/2006	Julia	7,478,036	B2	1/2009	Shen
7,043,425	B2	5/2006	Pao	7,487,088	B1	2/2009	Gorin
7,054,817	B2	5/2006	Shao	7,487,110	B2	2/2009	Bennett
7,058,890	B2	6/2006	George	7,493,259	B2	2/2009	Jones
7,062,488	B1	6/2006	Reisman	7,493,559	B1	2/2009	Wolff
7,069,220	B2	6/2006	Coffman	7,502,672	B1	3/2009	Kolls
7,072,834	B2	7/2006	Zhou	7,502,730	B2	3/2009	Wang
7,072,888	B1	7/2006	Perkins	7,502,738	B2	3/2009	Kennewick
7,076,362	B2	7/2006	Ohtsuji	7,512,906	B1	3/2009	Baier
7,082,469	B2	7/2006	Gold	7,516,076	B2	4/2009	Walker
7,085,708	B2	8/2006	Manson	7,529,675	B2	5/2009	Maes
7,092,928	B1	8/2006	Elad	7,536,297	B2	5/2009	Byrd
7,107,210	B2	9/2006	Deng	7,536,374	B2	5/2009	Au
7,107,218	B1	9/2006	Preston	7,542,894	B2	6/2009	Murata
7,110,951	B1	9/2006	Lemelson	7,546,382	B2	6/2009	Healey
7,127,395	B1	10/2006	Gorin	7,548,491	B2	6/2009	Macfarlane
7,127,400	B2	10/2006	Koch	7,552,054	B1	6/2009	Stifelman
7,130,390	B2	10/2006	Abburri	7,558,730	B2	7/2009	Davis
7,136,875	B2	11/2006	Anderson	7,574,362	B2	8/2009	Walker
				7,577,244	B2	8/2009	Taschereau
				7,606,708	B2	10/2009	Hwang
				7,606,712	B1	10/2009	Smith
				7,620,549	B2	11/2009	DiCristo

US 11,080,758 B2

Page 5

(56)	References Cited					
	U.S. PATENT DOCUMENTS					
	7,634,409	B2	12/2009	Kennewick	8,340,975	B1 12/2012 Rosenberger
	7,640,006	B2	12/2009	Portman	8,346,563	B1 1/2013 Hjelm
	7,640,160	B2	12/2009	DiCristo	8,370,147	B2 2/2013 Kennewick
	7,640,272	B2	12/2009	Mahajan	8,447,607	B2 5/2013 Weider
	7,672,931	B2	3/2010	Hurst-Hiller	8,447,651	B1 5/2013 Scholl
	7,676,365	B2	3/2010	Hwang	8,452,598	B2 5/2013 Kennewick
	7,676,369	B2	3/2010	Fujimoto	8,503,995	B2 8/2013 Ramer
	7,684,977	B2	3/2010	Morikawa	8,509,403	B2 8/2013 Chiu
	7,693,720	B2	4/2010	Kennewick	8,515,765	B2 8/2013 Baldwin
	7,697,673	B2	4/2010	Chiu	8,527,274	B2 9/2013 Freeman
	7,706,616	B2	4/2010	Kristensson	8,577,671	B1 11/2013 Barve
	7,729,913	B1	6/2010	Lee	8,589,161	B2 11/2013 Kennewick
	7,729,916	B2	6/2010	Coffman	8,612,205	B2 12/2013 Hanneman
	7,729,918	B2	6/2010	Walker	8,612,206	B2 12/2013 Chalabi
	7,729,920	B2	6/2010	Chaar	8,620,659	B2 12/2013 DiCristo
	7,734,287	B2	6/2010	Ying	8,719,005	B1 5/2014 Lee
	7,748,021	B2	6/2010	Obradovich	8,719,009	B2 5/2014 Baldwin
	7,774,333	B2	8/2010	Colledge	8,719,026	B2 5/2014 Kennewick
	7,788,084	B2	8/2010	Brun	8,731,929	B2 5/2014 Kennewick
	7,792,257	B1	9/2010	Vanier	8,738,380	B2 5/2014 Baldwin
	7,801,731	B2	9/2010	Odinak	8,849,652	B2 9/2014 Weider
	7,809,570	B2	10/2010	Kennewick	8,849,670	B2 9/2014 DiCristo
	7,818,176	B2	10/2010	Freeman	8,849,696	B2 9/2014 Pansari
	7,831,426	B2	11/2010	Bennett	8,849,791	B1 9/2014 Hertschuh
	7,831,433	B1	11/2010	Belvin	8,886,536	B2 11/2014 Freeman
	7,856,358	B2	12/2010	Ho	8,972,243	B1 3/2015 Strom
	7,873,519	B2	1/2011	Bennett	8,972,354	B1 3/2015 Telang
	7,873,523	B2	1/2011	Potter	8,983,839	B2 3/2015 Kennewick
	7,873,654	B2	1/2011	Bernard	9,009,046	B1 4/2015 Stewart
	7,881,936	B2	2/2011	Longe	9,015,049	B2 4/2015 Baldwin
	7,890,324	B2	2/2011	Bangalore	9,037,455	B1 5/2015 Faaborg
	7,894,849	B2	2/2011	Kass	9,070,366	B1 6/2015 Mathias
	7,902,969	B2	3/2011	Obradovich	9,070,367	B1 6/2015 Hoffmeister
	7,917,367	B2	3/2011	DiCristo	9,105,266	B2 8/2015 Baldwin
	7,920,682	B2	4/2011	Byrne	9,171,541	B2 10/2015 Kennewick
	7,949,529	B2	5/2011	Weider	9,269,097	B2 2/2016 Freeman
	7,949,537	B2	5/2011	Walker	9,305,548	B2 4/2016 Kennewick
	7,953,732	B2	5/2011	Frank	9,308,445	B1 4/2016 Merzenich
	7,974,875	B1	7/2011	Quilici	9,318,108	B2 4/2016 Gruber
	7,983,917	B2	7/2011	Kennewick	9,406,078	B2 8/2016 Freeman
	7,984,287	B2	7/2011	Gopalakrishnan	9,443,514	B1 9/2016 Taubman
	8,005,683	B2	8/2011	Tessel	9,502,025	B2 11/2016 Kennewick
	8,015,006	B2	9/2011	Kennewick	9,626,703	B2 4/2017 Kennewick, Sr.
	8,024,186	B1	9/2011	De Bonet	9,711,143	B2 7/2017 Kennewick
	8,027,965	B2	9/2011	Takehara	10,089,984	B2 10/2018 Kennewick
	8,032,383	B1	10/2011	Bhardwaj	2001/0011336	A1 8/2001 Sitka
	8,060,367	B2	11/2011	Keaveney	2001/0014857	A1 8/2001 Wang
	8,069,046	B2	11/2011	Kennewick	2001/0021905	A1 9/2001 Burnett
	8,073,681	B2	12/2011	Baldwin	2001/0039492	A1 11/2001 Nemoto
	8,077,975	B2	12/2011	Ma	2001/0041980	A1 11/2001 Howard
	8,078,502	B2	12/2011	Hao	2001/0047261	A1 11/2001 Kassan
	8,082,153	B2	12/2011	Coffman	2001/0049601	A1 12/2001 Kroeker
	8,086,463	B2	12/2011	Ativanichayaphong	2001/0054087	A1 12/2001 Flom
	8,103,510	B2	1/2012	Sato	2002/0002548	A1 1/2002 Roundtree
	8,112,275	B2	2/2012	Kennewick	2002/0007267	A1 1/2002 Batchilo
	8,140,327	B2	3/2012	Kennewick	2002/0010584	A1 1/2002 Schultz
	8,140,335	B2	3/2012	Kennewick	2002/0015500	A1 2/2002 Belt
	8,145,489	B2	3/2012	Freeman	2002/0022927	A1 2/2002 Lemelson
	8,150,694	B2	4/2012	Kennewick	2002/0022956	A1 2/2002 Ukrainczyk
	8,155,962	B2	4/2012	Kennewick	2002/0029186	A1 3/2002 Roth
	8,170,867	B2	5/2012	Germain	2002/0029261	A1 3/2002 Shibata
	8,180,037	B1	5/2012	Delker	2002/0032752	A1 3/2002 Gold
	8,190,436	B2	5/2012	Bangalore	2002/0035501	A1 3/2002 Handel
	8,195,468	B2	6/2012	Weider	2002/0040297	A1 4/2002 Tsiao
	8,200,485	B1	6/2012	Lee	2002/0049535	A1 4/2002 Rigo
	8,204,751	B1	6/2012	Di Fabbizio	2002/0049805	A1 4/2002 Yamada
	8,219,399	B2	7/2012	Lutz	2002/0059068	A1 5/2002 Rose
	8,219,599	B2	7/2012	Tunstall-Pedoe	2002/0065568	A1 5/2002 Silfvast
	8,224,652	B2	7/2012	Wang	2002/0067839	A1 6/2002 Heinrich
	8,255,224	B2	8/2012	Singleton	2002/0069059	A1 6/2002 Smith
	8,326,599	B2	12/2012	Tomeh	2002/0069071	A1 6/2002 Knockeart
	8,326,627	B2	12/2012	Kennewick	2002/0073176	A1 6/2002 Ikeda
	8,326,634	B2	12/2012	DiCristo	2002/0082911	A1 6/2002 Dunn
	8,326,637	B2	12/2012	Baldwin	2002/0087312	A1 7/2002 Lee
	8,332,224	B2	12/2012	DiCristo	2002/0087326	A1 7/2002 Lee
					2002/0087525	A1 7/2002 Abbott
					2002/0107694	A1 8/2002 Lerg
					2002/0120609	A1 8/2002 Lang
					2002/0124050	A1 9/2002 Middeljans

US 11,080,758 B2

Page 6

(56)	References Cited			2004/0166832	A1	8/2004	Portman
	U.S. PATENT DOCUMENTS			2004/0167771	A1	8/2004	Duan
				2004/0172247	A1	9/2004	Yoon
				2004/0172258	A1	9/2004	Dominach
2002/0133347	A1	9/2002	Schoneburg	2004/0189697	A1	9/2004	Fukuoka
2002/0133354	A1	9/2002	Ross	2004/0193408	A1	9/2004	Hunt
2002/0133402	A1	9/2002	Faber	2004/0193420	A1	9/2004	Kennewick
2002/0135618	A1	9/2002	Maes	2004/0199375	A1	10/2004	Ehsani
2002/0138248	A1	9/2002	Corston-Oliver	2004/0199387	A1	10/2004	Wang
2002/0143532	A1	10/2002	McLean	2004/0199389	A1	10/2004	Geiger
2002/0143535	A1	10/2002	Kist	2004/0201607	A1	10/2004	Mulvey
2002/0152260	A1	10/2002	Chen	2004/0204939	A1	10/2004	Liu
2002/0161646	A1	10/2002	Galley	2004/0205671	A1	10/2004	Sukehiro
2002/0161647	A1	10/2002	Gailey	2004/0220800	A1	11/2004	Kong
2002/0169597	A1	11/2002	Fain	2004/0243393	A1	12/2004	Wang
2002/0173333	A1	11/2002	Buchholz	2004/0243417	A9	12/2004	Pitts
2002/0173961	A1	11/2002	Guerra	2004/0247092	A1	12/2004	Timmins
2002/0178003	A1	11/2002	Gehrke	2004/0249636	A1	12/2004	Applebaum
2002/0184373	A1	12/2002	Maes	2005/0015256	A1	1/2005	Kargman
2002/0188455	A1	12/2002	Shioda	2005/0021331	A1	1/2005	Huang
2002/0188602	A1	12/2002	Stubler	2005/0021334	A1	1/2005	Iwahashi
2002/0198714	A1	12/2002	Zhou	2005/0021470	A1	1/2005	Martin
2003/0005033	A1	1/2003	Mohan	2005/0021826	A1	1/2005	Kumar
2003/0014261	A1	1/2003	Kageyama	2005/0033574	A1	2/2005	Kim
2003/0016835	A1	1/2003	Elko	2005/0033582	A1	2/2005	Gadd
2003/0036903	A1	2/2003	Konopka	2005/0043940	A1	2/2005	Elder
2003/0046071	A1	3/2003	Wyman	2005/0043956	A1	2/2005	Aoyama
2003/0046281	A1	3/2003	Son	2005/0060142	A1	3/2005	Visser
2003/0046346	A1	3/2003	Mumick	2005/0080632	A1	4/2005	Endo
2003/0064709	A1	4/2003	Gailey	2005/0080821	A1	4/2005	Breil
2003/0065427	A1	4/2003	Funk	2005/0102282	A1	5/2005	Linden
2003/0069734	A1	4/2003	Everhart	2005/0114116	A1	5/2005	Fiedler
2003/0069880	A1	4/2003	Harrison	2005/0125232	A1	6/2005	Gadd
2003/0088421	A1	5/2003	Maes	2005/0131673	A1	6/2005	Koizumi
2003/0093419	A1	5/2003	Bangalore	2005/0137850	A1	6/2005	Odell
2003/0097249	A1	5/2003	Walker	2005/0137877	A1	6/2005	Oesterling
2003/0110037	A1	6/2003	Walker	2005/0138168	A1	6/2005	Hoffman
2003/0112267	A1	6/2003	Belrose	2005/0143994	A1	6/2005	Mori
2003/0115062	A1	6/2003	Walker	2005/0144013	A1	6/2005	Fujimoto
2003/0120493	A1	6/2003	Gupta	2005/0144187	A1	6/2005	Che
2003/0135488	A1	7/2003	Amir	2005/0149319	A1	7/2005	Honda
2003/0144846	A1	7/2003	Denenberg	2005/0169441	A1	8/2005	Yacoub
2003/0154082	A1	8/2003	Toguri	2005/0216254	A1	9/2005	Gupta
2003/0158731	A1	8/2003	Falcon	2005/0222763	A1	10/2005	Uyeki
2003/0161448	A1	8/2003	Parolkar	2005/0234637	A1	10/2005	Obradovich
2003/0167167	A1	9/2003	Gong	2005/0234727	A1	10/2005	Chiu
2003/0174155	A1	9/2003	Weng	2005/0240412	A1	10/2005	Fujita
2003/0182123	A1	9/2003	Mitsuyoshi	2005/0246174	A1	11/2005	DeGolia
2003/0182132	A1	9/2003	Niemoeller	2005/0283364	A1	12/2005	Longe
2003/0187643	A1	10/2003	VanThong	2005/0283532	A1	12/2005	Kim
2003/0204492	A1	10/2003	Wolf	2005/0283752	A1	12/2005	Fruchter
2003/0206640	A1	11/2003	Malvar	2006/0041431	A1	2/2006	Maes
2003/0212550	A1	11/2003	Ubale	2006/0046740	A1	3/2006	Johnson
2003/0212558	A1	11/2003	Matula	2006/0047509	A1	3/2006	Ding
2003/0212562	A1	11/2003	Patel	2006/0072738	A1	4/2006	Louis
2003/0216919	A1	11/2003	Roushar	2006/0074670	A1	4/2006	Weng
2003/0225825	A1	12/2003	Healey	2006/0074671	A1	4/2006	Farmaner
2003/0233230	A1	12/2003	Ammicht	2006/0080098	A1	4/2006	Campbell
2003/0236664	A1	12/2003	Sharma	2006/0100851	A1	5/2006	Schonebeck
2004/0006475	A1	1/2004	Ehlen	2006/0106769	A1	5/2006	Gibbs
2004/0010358	A1	1/2004	Oesterling	2006/0129409	A1	6/2006	Mizutani
2004/0025115	A1	2/2004	Sienel	2006/0130002	A1	6/2006	Hirayama
2004/0030741	A1	2/2004	Wolton	2006/0149633	A1	7/2006	Voisin
2004/0036601	A1	2/2004	Obradovich	2006/0182085	A1	8/2006	Sweeney
2004/0044516	A1	3/2004	Kennewick	2006/0206310	A1	9/2006	Ravikumar
2004/0093567	A1	5/2004	Schabes	2006/0217133	A1	9/2006	Christenson
2004/0098245	A1	5/2004	Walker	2006/0236343	A1	10/2006	Chang
2004/0102977	A1	5/2004	Metzler	2006/0242017	A1	10/2006	Libes
2004/0117179	A1	6/2004	Balasuriya	2006/0247919	A1	11/2006	Specht
2004/0117804	A1	6/2004	Scahill	2006/0253247	A1	11/2006	de Silva
2004/0122673	A1	6/2004	Park	2006/0253281	A1	11/2006	Letzt
2004/0122674	A1	6/2004	Bangalore	2006/0285662	A1	12/2006	Yin
2004/0128514	A1	7/2004	Rhoads	2007/0011159	A1	1/2007	Hillis
2004/0133793	A1	7/2004	Ginter	2007/0033005	A1	2/2007	Di Cristo
2004/0140989	A1	7/2004	Papageorge	2007/0033020	A1	2/2007	Francois
2004/0143440	A1	7/2004	Prasad	2007/0033526	A1	2/2007	Thompson
2004/0148154	A1	7/2004	Acero	2007/0038436	A1	2/2007	Cristo
2004/0148170	A1	7/2004	Acero	2007/0038445	A1	2/2007	Helbing
2004/0158555	A1	8/2004	Seedman	2007/0043569	A1	2/2007	Potter

US 11,080,758 B2

Page 7

(56)

References Cited

U.S. PATENT DOCUMENTS

2007/0043574	A1	2/2007	Coffman	2008/0294994	A1	11/2008	Kruger
2007/0043868	A1	2/2007	Kumar	2008/0306743	A1	12/2008	Di Fabbriozio
2007/0050191	A1	3/2007	Weider	2008/0319751	A1	12/2008	Kennewick
2007/0050279	A1	3/2007	Huang	2009/0006077	A1	1/2009	Keaveney
2007/0055525	A1	3/2007	Kennewick	2009/0006194	A1	1/2009	Prashantidharan
2007/0060114	A1	3/2007	Ramer	2009/0018829	A1	1/2009	Kuperstein
2007/0061067	A1	3/2007	Zeinstra	2009/0024476	A1	1/2009	Baar
2007/0061735	A1	3/2007	Hoffberg	2009/0030686	A1	1/2009	Weng
2007/0067310	A1	3/2007	Gupta	2009/0052635	A1	2/2009	Jones
2007/0073544	A1	3/2007	Millett	2009/0055176	A1	2/2009	Hu
2007/0078708	A1	4/2007	Yu	2009/0067599	A1	3/2009	Agarwal
2007/0078709	A1	4/2007	Rajaram	2009/0076827	A1	3/2009	Bulitta
2007/0078814	A1	4/2007	Flowers	2009/0106029	A1	4/2009	DeLine
2007/0094003	A1	4/2007	Huang	2009/0117885	A1	5/2009	Roth
2007/0100797	A1	5/2007	Thun	2009/0144131	A1	6/2009	Chiu
2007/0106499	A1	5/2007	Dahlgren	2009/0144271	A1	6/2009	Richardson
2007/0112555	A1	5/2007	Lavi	2009/0150156	A1	6/2009	Kennewick
2007/0112630	A1	5/2007	Lau	2009/0157382	A1	6/2009	Bar
2007/0118357	A1	5/2007	Kasravi	2009/0164216	A1	6/2009	Chengalvarayan
2007/0124057	A1	5/2007	Prieto	2009/0171664	A1	7/2009	Kennewick
2007/0135101	A1	6/2007	Ramati	2009/0171912	A1	7/2009	Nash
2007/0146833	A1	6/2007	Satomi	2009/0197582	A1	8/2009	Lewis
2007/0162296	A1	7/2007	Altberg	2009/0216540	A1	8/2009	Tessel
2007/0174258	A1*	7/2007	Jones	2009/0248565	A1	10/2009	Chuang
2007/0179778	A1	8/2007	Gong	2009/0248605	A1	10/2009	Mitchell
2007/0185859	A1	8/2007	Flowers	2009/0259561	A1	10/2009	Boys
2007/0186165	A1	8/2007	Maislos	2009/0259646	A1	10/2009	Fujita
2007/0192309	A1	8/2007	Fischer	2009/0265163	A1	10/2009	Li
2007/0198267	A1	8/2007	Jones	2009/0271194	A1	10/2009	Davis
2007/0203699	A1	8/2007	Nagashima	2009/0273563	A1	11/2009	Pryor
2007/0203736	A1	8/2007	Ashton	2009/0276700	A1	11/2009	Anderson
2007/0208732	A1	9/2007	Flowers	2009/0287680	A1	11/2009	Paek
2007/0214182	A1	9/2007	Rosenberg	2009/0299745	A1	12/2009	Kennewick
2007/0250901	A1	10/2007	McIntire	2009/0299857	A1	12/2009	Brubaker
2007/0265850	A1	11/2007	Kennewick	2009/0304161	A1	12/2009	Pettyjohn
2007/0266257	A1	11/2007	Camaisa	2009/0307031	A1	12/2009	Winkler
2007/0276651	A1	11/2007	Bliss	2009/0313026	A1	12/2009	Coffman
2007/0294615	A1	12/2007	Sathe	2009/0319517	A1	12/2009	Guha
2007/0299824	A1	12/2007	Pan	2010/0023320	A1	1/2010	Cristo
2008/0010135	A1	1/2008	Schrock	2010/0023331	A1	1/2010	Duta
2008/0014908	A1	1/2008	Vasant	2010/0029261	A1	2/2010	Mikkelsen
2008/0032622	A1	2/2008	Kopra	2010/0036967	A1	2/2010	Caine
2008/0034032	A1	2/2008	Healey	2010/0049501	A1	2/2010	Kennewick
2008/0046311	A1	2/2008	Shahine	2010/0049514	A1	2/2010	Kennewick
2008/0059188	A1	3/2008	Konopka	2010/0057443	A1	3/2010	Cristo
2008/0065386	A1	3/2008	Cross	2010/0063880	A1	3/2010	Atsmon
2008/0065389	A1	3/2008	Cross	2010/0064025	A1	3/2010	Nelimarkka
2008/0065390	A1	3/2008	Ativanichayaphong	2010/0076778	A1	3/2010	Kondrk
2008/0086455	A1	4/2008	Meisels	2010/0094707	A1	4/2010	Freer
2008/0091406	A1	4/2008	Baldwin	2010/0138300	A1	6/2010	Wallis
2008/0103761	A1	5/2008	Printz	2010/0145700	A1	6/2010	Kennewick
2008/0103781	A1	5/2008	Wasson	2010/0185512	A1	7/2010	Borger
2008/0104071	A1	5/2008	Pragada	2010/0191856	A1	7/2010	Gupta
2008/0109285	A1	5/2008	Reuther	2010/0204986	A1	8/2010	Kennewick
2008/0115163	A1	5/2008	Gilboa	2010/0204994	A1	8/2010	Kennewick
2008/0126091	A1	5/2008	Clark	2010/0217604	A1	8/2010	Baldwin
2008/0126284	A1	5/2008	Forbes	2010/0268536	A1	10/2010	Suendermann
2008/0133215	A1	6/2008	Sarukkai	2010/0286985	A1	11/2010	Kennewick
2008/0140385	A1	6/2008	Mahajan	2010/0299142	A1	11/2010	Freeman
2008/0147396	A1	6/2008	Wang	2010/0312547	A1	12/2010	Van Os
2008/0147410	A1	6/2008	Odinak	2010/0312566	A1	12/2010	Odinak
2008/0147637	A1	6/2008	Li	2010/0318357	A1	12/2010	Istvan
2008/0154604	A1	6/2008	Sathish	2010/0331064	A1	12/2010	Michelstein
2008/0162471	A1	7/2008	Bernard	2011/0022393	A1	1/2011	Waller
2008/0177530	A1	7/2008	Cross	2011/0106527	A1	5/2011	Chiu
2008/0184164	A1	7/2008	Di Fabbriozio	2011/0112827	A1	5/2011	Kennewick
2008/0189110	A1	8/2008	Freeman	2011/0112921	A1	5/2011	Kennewick
2008/0189187	A1	8/2008	Hao	2011/0119049	A1	5/2011	Ylonen
2008/0228496	A1	9/2008	Yu	2011/0131036	A1	6/2011	DiCristo
2008/0235023	A1	9/2008	Kennewick	2011/0131045	A1	6/2011	Cristo
2008/0235027	A1	9/2008	Cross	2011/0231182	A1	9/2011	Weider
2008/0269958	A1	10/2008	Filev	2011/0231188	A1	9/2011	Kennewick
2008/0270135	A1	10/2008	Goel	2011/0238409	A1	9/2011	Larcheveque
2008/0270224	A1	10/2008	Portman	2011/0307167	A1	12/2011	Taschereau
2008/0294437	A1	11/2008	Nakano	2012/0022857	A1	1/2012	Baldwin
				2012/0041753	A1	2/2012	Dymetman
				2012/0046935	A1	2/2012	Nagao
				2012/0101809	A1	4/2012	Kennewick
				2012/0101810	A1	4/2012	Kennewick

US 11,080,758 B2

Page 8

(56) **References Cited**
U.S. PATENT DOCUMENTS

2012/0109753 A1 5/2012 Kennewick
 2012/0150620 A1 6/2012 Mandyam
 2012/0150636 A1 6/2012 Freeman
 2012/0239498 A1 9/2012 Ramer
 2012/0240060 A1 9/2012 Pennington
 2012/0265528 A1 10/2012 Gruber
 2012/0278073 A1 11/2012 Weider
 2013/0006734 A1 1/2013 Ocko
 2013/0054228 A1 2/2013 Baldwin
 2013/0060625 A1 3/2013 Davis
 2013/0080177 A1 3/2013 Chen
 2013/0211710 A1 8/2013 Kennewick
 2013/0253929 A1 9/2013 Weider
 2013/0254314 A1 9/2013 Chow
 2013/0297293 A1 11/2013 Cristo
 2013/0304473 A1 11/2013 Baldwin
 2013/0311324 A1 11/2013 Stoll
 2013/0332454 A1 12/2013 Stuhec
 2013/0339022 A1 12/2013 Baldwin
 2014/0006951 A1 1/2014 Hunter
 2014/0012577 A1 1/2014 Freeman
 2014/0025371 A1 1/2014 Min
 2014/0108013 A1 4/2014 Cristo
 2014/0136187 A1 5/2014 Wolverton
 2014/0136259 A1 5/2014 Kinsey, II
 2014/0156278 A1 6/2014 Kennewick
 2014/0195238 A1 7/2014 Terao
 2014/0236575 A1 8/2014 Tur
 2014/0249821 A1 9/2014 Kennewick
 2014/0249822 A1 9/2014 Baldwin
 2014/0278413 A1 9/2014 Pitschel
 2014/0278416 A1 9/2014 Schuster
 2014/0288934 A1 9/2014 Kennewick
 2014/0330552 A1 11/2014 Bangalore
 2014/0337007 A1 11/2014 Waibel
 2014/0365222 A1 12/2014 Weider
 2015/0019211 A1 1/2015 Simard
 2015/0019217 A1 1/2015 Cristo
 2015/0019227 A1 1/2015 Anandarajah
 2015/0066479 A1 3/2015 Pasupalak
 2015/0066627 A1 3/2015 Freeman
 2015/0073910 A1 3/2015 Kennewick
 2015/0095159 A1 4/2015 Kennewick
 2015/0142447 A1 5/2015 Kennewick
 2015/0170641 A1 6/2015 Kennewick
 2015/0193379 A1 7/2015 Mehta
 2015/0199339 A1 7/2015 Mirkin
 2015/0228276 A1 8/2015 Baldwin
 2015/0293917 A1 10/2015 Bufo
 2015/0348544 A1 12/2015 Baldwin
 2015/0348551 A1 12/2015 Gruber
 2015/0364133 A1 12/2015 Freeman
 2016/0049152 A1 2/2016 Kennewick
 2016/0078482 A1 3/2016 Kennewick
 2016/0078491 A1 3/2016 Kennewick
 2016/0078504 A1 3/2016 Kennewick
 2016/0078773 A1 3/2016 Carter
 2016/0110347 A1 4/2016 Kennewick
 2016/0148610 A1 5/2016 Kennewick
 2016/0148612 A1 5/2016 Guo
 2016/0188292 A1 6/2016 Carter
 2016/0188573 A1 6/2016 Tang
 2016/0217785 A1 7/2016 Kennewick
 2016/0335676 A1 11/2016 Freeman
 2017/0004588 A1 1/2017 Isaacson

FOREIGN PATENT DOCUMENTS

CN 101236635 A 8/2008
 CN 103229206 A 7/2013
 EP 1320043 A2 6/2003
 EP 1646037 4/2006
 FR 2867583 A1 * 9/2005 G10L 15/26
 JP H08263258 10/1996
 JP H11249773 9/1999

JP 2001071289 3/2001
 JP 2002297626 10/2002
 JP 2006146881 6/2006
 JP 2008027454 2/2008
 JP 2008058465 3/2008
 JP 2008139928 6/2008
 JP 2011504304 2/2011
 JP 2012518847 8/2012
 WO 9946763 9/1999
 WO 0021232 4/2000
 WO 0046792 8/2000
 WO 0129742 4/2001
 WO 0129742 A2 4/2001
 WO 0171609 A2 9/2001
 WO 0178065 10/2001
 WO 2004072954 8/2004
 WO 2005010702 A2 2/2005
 WO 2007019318 1/2007
 WO 2007021587 1/2007
 WO 2007027546 1/2007
 WO 2007027989 1/2007
 WO 2008098039 1/2008
 WO 2008118195 1/2008
 WO 2009075912 1/2009
 WO 2009145796 1/2009
 WO 2009111721 9/2009
 WO 2010096752 1/2010
 WO 2016044290 3/2016
 WO 2016044319 3/2016
 WO 2016044321 3/2016
 WO M16044316 3/2016
 WO 2016061309 4/2016

OTHER PUBLICATIONS

Arrington, Michael, "Google Redefines GPS Navigation Landscape: Google Maps Navigation for Android 2.0", TechCrunch, printed from the Internet <<http://www.techcrunch.com/2009/10/28/google-redefines-car-gps-navigation-google-maps-navigation-android/>>, Oct. 28, 2009, 4 pages.

Bazzi, Issam et al., "Heterogeneous Lexical Units for Automatic Speech Recognition: Preliminary Investigations", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, Jun. 5-9, 2000, XP010507574, pp. 1257-1260.

Belvin, Robert, et al., "Development of the HRL Route Navigation Dialogue System", Proceedings of the First International Conference on Human Language Technology Research, San Diego, 2001, pp. 1-5.

Chai et al., "MIND: A Semantics-Based Multimodal Interpretation Framework for Conversational Systems", Proceedings of the International Class Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Jun. 2002, pp. 37-46.

Cheyen et al., "Multimodal Maps: An Agent-Based Approach", International Conference on Cooperative Multimodal Communication (CMC/95), May 24-26, 1995, pp. 111-121.

Davis, Z., et al., A Personal Handheld Multi-Modal Shopping Assistant, IEEE, 2006, 9 pages.

El Meliani et al., "A Syllabic-Filler-Based Continuous Speech Recognizer for Unlimited Vocabulary", Canadian conference on Electrical and Computer Engineering, vol. 2, Sep. 5-8, 1995, pp. 1007-1010.

Elio et al., "On Abstract Task Models and Conversation Policies" in Workshop on Specifying and Implementing Conversation Policies, Autonomous Agents '99, Seattle, 1999, 10 pages.

Kirchhoff, Katrin, "Syllable-Level Desynchronisation of Phonetic Features for Speech Recognition", Proceedings of the Fourth International Conference on Spoken Language, 1996, ICSLP 96, vol. 4, IEEE, 1996, 3 pages.

Kuhn, Thomas, et al., "Hybrid In-Car Speech Recognition for Mobile Multimedia Applications", Vehicular Technology Conference, IEEE, Jul. 1999, pp. 2009-2013.

Lin, Bor-shen, et al., "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History", ASRU'99, 1999, 4 pages.

US 11,080,758 B2

Page 9

(56)

References Cited

OTHER PUBLICATIONS

Lind, R., et al., The Network Vehicle—A Glimpse into the Future of Mobile Multi-Media, IEEE Aerosp. Electron. Systems Magazine, vol. 14, No. 9, Sep. 1999, pp. 27-32.

Mao, Mark Z., "Automatic Training Set Segmentation for Multi-Pass Speech Recognition", Department of Electrical Engineering, Stanford University, CA, copyright 2005, IEEE, pp. I-685 to I-688.

O'Shaughnessy, Douglas, "Interacting with Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, vol. 91, No. 9, Sep. 1, 2003, XP011100665, pp. 1272-1305.

Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., Jul. 28, 2002, 1 page.

Turunen, "Adaptive Interaction Methods in Speech User Interfaces", Conference on Human Factors in Computing Systems, Seattle, Washington, 2001, pp. 91-92.

Vanhoeck, Vincent, "Confidence Scoring and Rejection Using Multi-Pass Speech Recognition", Nuance Communications, Menlo Park, CA, 2005, 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, 1998, 4 pages.

Wu, Su-Lin, et al., "Incorporating Information from Syllable-Length Time Scales into Automatic Speech Recognition", Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, vol. 2 IEEE, 1998, 4 pages.

Wu, Su-Lin, et al., "Integrating Syllable Boundary Information into Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, 1997, vol. 2, IEEE, 1997, 4 pages.

Zhao, Yilin, "Telematics: Safe and Fun Driving", IEEE Intelligent Systems, vol. 17, Issue 1, 2002, pp. 10-14.

Office Action issued in Chinese Patent Application No. 201580060519.1 dated Mar. 13, 2020, with its English translation, 40 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01367, Decision Granting Institution of Inter Partes Review of U.S. Pat. No. 8,073,681 B2, entered Mar. 11, 2021, 22 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01390, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 7,818,176, entered Mar. 11, 2021, 27 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-

01346, Decision Granting Institution of Inter Partes Review of U.S. Pat. No. 9,015,049 B2, entered Feb. 4, 2021, 24 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01374, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 8,886,536 B2, entered Feb. 4, 2021, 37 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01377, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 8,886,536 B2, entered Feb. 4, 2021, 29 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01380, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 9,626,703 B2, entered Feb. 4, 2021, 28 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01381, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 9,626,703 B2, entered Feb. 4, 2021, 29 pages.

Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., AZZ Development Center, Inc. d/b/a/ Lab126, Rawles LLC, Amzn Mobile LLC, Amzn Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) v. VB Assets, LLC, IPR2020-01388, Decision Denying Institution of Inter Partes Review of U.S. Pat. No. 9,269,097 B2, entered Feb. 24, 2021, 25 pages.

Asthana, A., et al., "A Small Domain Communications System for Personalized Shopping Assistance", Proceedings of 1994 International Conference on Personal Wireless Communications, IEEE Press, Aug. 1994, pp. 199-203.

Huang et al., "Spoken Language Processing: A Guide to Theory, Algorithm, and System Development", Prentice Hall, 2001, 1010 pages.

IEEE 100 The Authoritative Dictionary of IEEE Standards Terms, Seventh Edition, Standards Information Network, IEEE Press, Print ISBN 0-7381-2601-2, Published Dec. 2000, 3 pages.

Lucente, Mark, "Conversational Interfaces for E-Commerce Applications", Communications of the ACM, vol. 43, No. 9, Sep. 2000, pp. 59-61.

Seneff, Stephanie, et al., "Hypothesis Selection and Resolution in the Mercury Flight Reservation System", Spoken Language Systems Group, MIT, 2001, 8 pages.

* cited by examiner

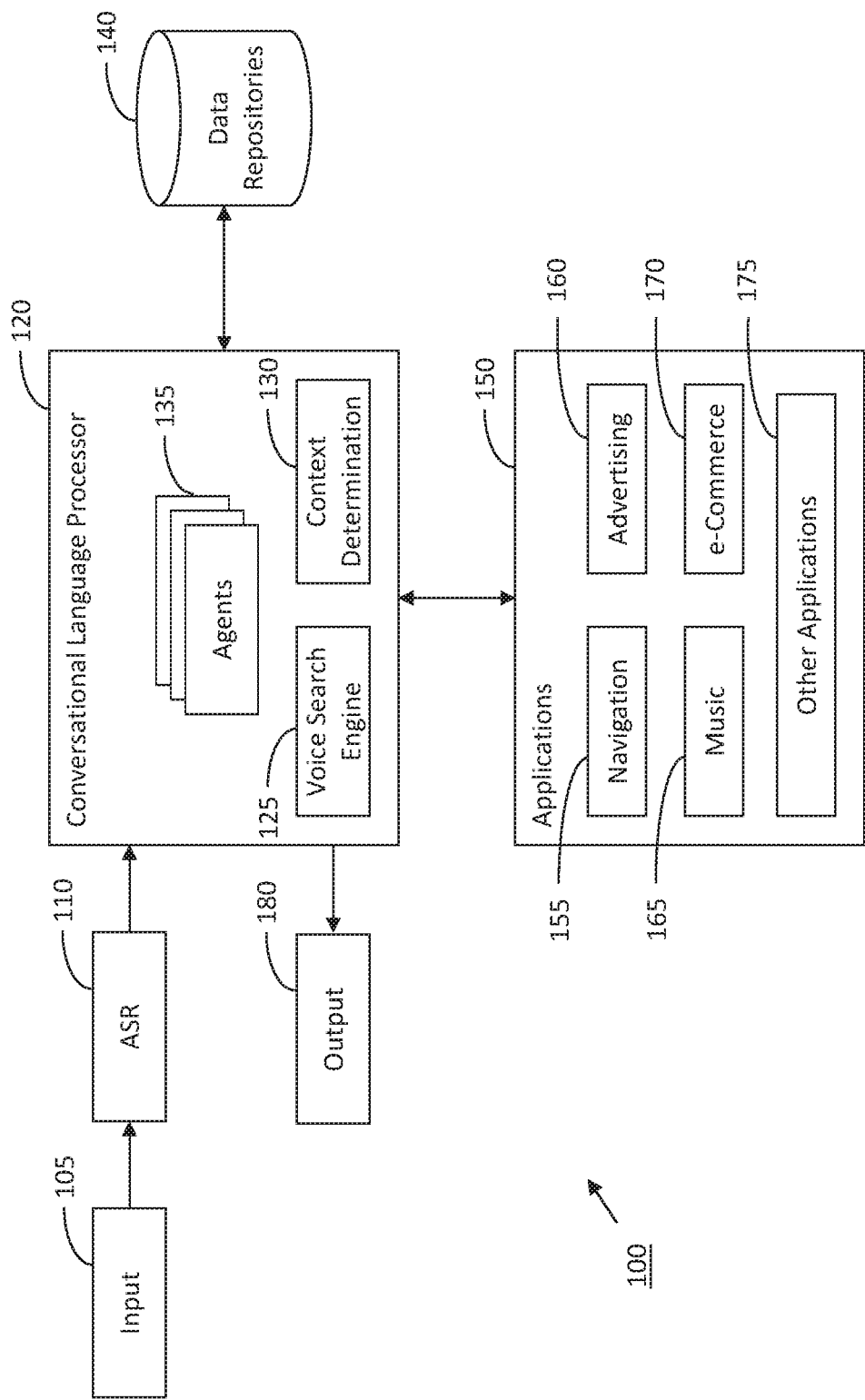


Figure 1

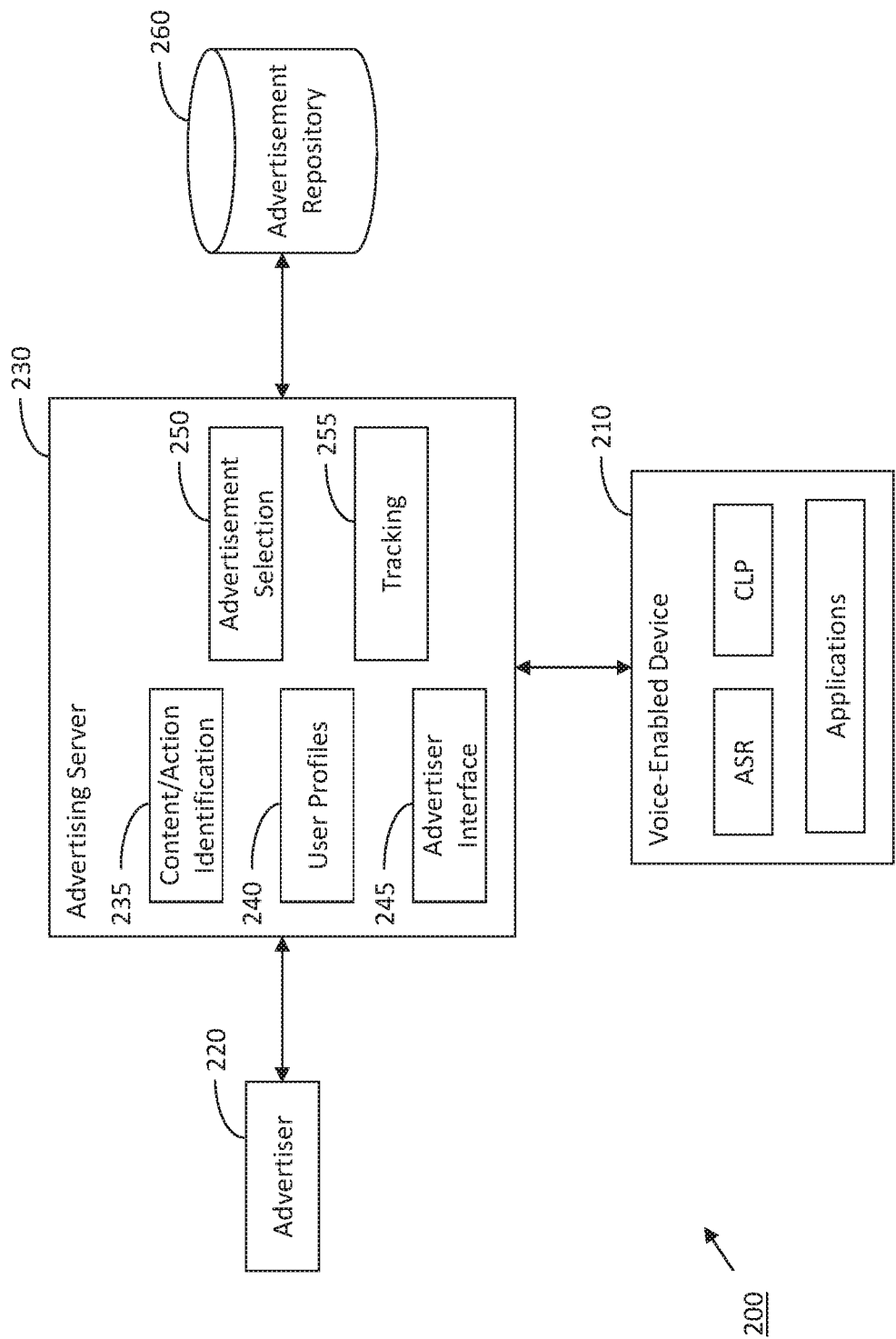


Figure 2

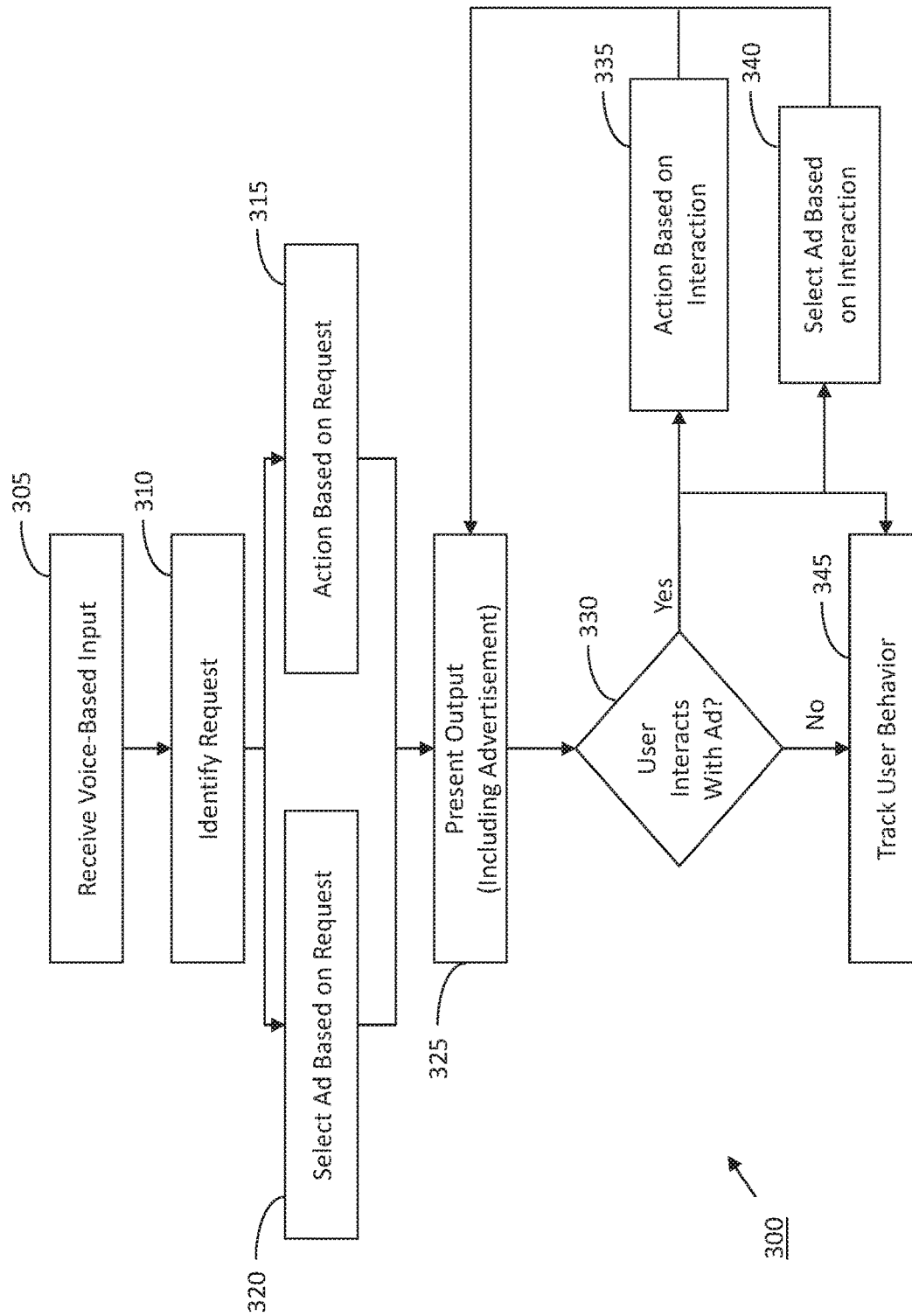


Figure 3

US 11,080,758 B2

1

SYSTEM AND METHOD FOR DELIVERING TARGETED ADVERTISEMENTS AND/OR PROVIDING NATURAL LANGUAGE PROCESSING BASED ON ADVERTISEMENTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. patent application Ser. No. 15/223,870, entitled "System and Method for Delivering Targeted Advertisements and/or Providing Natural Language Processing Based on Advertisements", filed Jul. 29, 2016 (which issued as U.S. Pat. No. 10,134,060 on Nov. 20, 2018), which is a continuation of U.S. patent application Ser. No. 14/836,606, entitled "System and Method for Delivering Targeted Advertisements and/or Providing Natural Language Processing Based on Advertisements," filed Aug. 26, 2015 (which issued as U.S. Pat. No. 9,406,078 on Aug. 2, 2016), which is a continuation of U.S. patent application Ser. No. 14/537,598, entitled "System and Method for Delivering Targeted Advertisements and/or Providing Natural Language Processing Based on Advertisements," filed Nov. 10, 2014 (which issued as U.S. Pat. No. 9,269,097 on Feb. 23, 2016), which is a continuation of U.S. patent application Ser. No. 14/016,757, entitled "System and Method for Delivering Targeted Advertisements and Tracking Advertisement Interactions in Voice Recognition Context," filed Sep. 3, 2013 (which issued as U.S. Pat. No. 8,886,536 on Nov. 11, 2014), which is a continuation of U.S. patent application Ser. No. 13/371,870, entitled "System and Method for Delivering Target Advertisements and Tracking Advertisement Interactions in Voice Recognition Context," filed Feb. 13, 2012 (which issued as U.S. Pat. No. 8,527,274 on Sep. 3, 2013), which is a continuation of U.S. patent application Ser. No. 12/847,564, entitled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input," filed Jul. 30, 2010 (which issued as U.S. Pat. No. 8,145,489 on Mar. 27, 2012), which is a continuation of U.S. patent application Ser. No. 11/671,526, entitled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input," filed Feb. 6, 2007 (which issued as U.S. Pat. No. 7,818,176 on Oct. 19, 2010), the contents of which are hereby incorporated by reference in their entirety.

FIELD OF THE INVENTION

The present invention relates to delivering targeted advertisements and/or processing natural language processing based on advertisements.

BACKGROUND OF THE INVENTION

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

2

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.

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.

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.

Existing systems suffer from these and other problems.

SUMMARY OF THE INVENTION

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.

According to various aspects of the invention, advertisers may create advertisements, which may be stored in an advertisement repository. For example, advertisements may include sponsored messages, calls to action, purchase opportunities, trial downloads, or any other marketing communi-

US 11,080,758 B2

3

cation, 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.

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

BRIEF DESCRIPTION OF THE DRAWINGS

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

FIG. 2 illustrates a block diagram of an exemplary advertising system according to various aspects of the invention.

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

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 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 U.S. patent application Ser. No. 10/452,147, entitled “Systems and Methods for Responding to Natural Language Speech Utterance,” filed Jun. 3, 2003, which issued as U.S. Pat. No. 7,398,209 on Jul. 8, 2008, and U.S. patent application Ser. No. 10/618,633, entitled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance,” filed Jun. 15, 2003, which issued as U.S. Pat. No. 7,693,720 on Apr. 6, 2010, both of which are hereby incorporated by reference in their entirety. For example, as described in U.S. patent application Ser. No. 10/452,147, the system **100** may include a speech recognition engine (e.g., an Automatic Speech Recognizer **110**) that may recognize words and phrases in an utterance using entries in one or more dictionary and phrase tables. In addition, as further described therein, fuzzy set possibilities or prior probabilities for the words in the dictionary and phrase tables may be dynamically updated to maximize the

4

probability of correct recognition at each stage of the dialog (e.g., the probabilities or possibilities may be dynamically updated based on application domains, questions or commands, contexts, user profiles and preferences, user dialog histories, recognizer dictionary and phrase tables, word spellings, and/or other criteria).

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 the 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 U.S. patent application Ser. No. 11/513,269, entitled “Dynamic Speech Sharpening,” filed Aug. 31, 2006, which issued as U.S. Pat. No. 7,634,409 on Dec. 15, 2009, and which is hereby incorporated by reference in its entirety.

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 **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).

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).

Conversational language processor **120** may build long-term and/or short-term shared knowledge in one or more knowledge sources. For example, shared knowledge sources may include information about previous utterances, requests, and other user interactions to inform generating an

US 11,080,758 B2

5

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 U.S. patent application Ser. No. 11/197,504, entitled “Systems and Methods for Responding to Natural Language Speech Utterance,” filed Aug. 5, 2005, which issued as U.S. Pat. No. 7,640,160 on Dec. 29, 2009, and U.S. patent application Ser. No. 11/212,693, entitled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions,” filed Aug. 29, 2005, which issued as U.S. Pat. No. 7,949,529 on May 24, 2011, both of which are hereby incorporated by reference in their entirety.

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 U.S. patent application Ser. No. 11/200,164, entitled “System and Method of Supporting Adaptive Misrecognition in Conversational Speech,” filed Aug. 10, 2005, which issued as U.S. Pat. No. 7,620,549 on Nov. 17, 2009, and U.S. patent application Ser. No. 11/580,926, entitled “System and Method for a Cooperative Conversational Voice User Interface,” filed Oct. 16, 2006, which issued as U.S. Pat. No. 8,073,681 on Dec. 6, 2011, both of which are hereby incorporated by reference in their entirety. For example, as described in U.S. patent application Ser. No. 11/200,164, an environmental model may be accessed to determine user location, user activity, track user actions, and/or other environmental information to invoke context, domain knowledge, preferences, and/or other cognitive qualities to enhance the interpretation of questions and/or commands. In addition, as further described therein, based on information received from a general cognitive model, the environmental model, and/or a personalized cognitive model, which provide statistical abstracts of user interaction patterns, the system **100** may enhance responses to commands and questions by including a prediction of user behavior.

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 down-

6

loads, 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.

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, 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.

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 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**.

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.

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 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,

US 11,080,758 B2

7

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).

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).

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. 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.

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.

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

8

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).

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 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 U.S. Patent Applications and U.S. Patents, which are hereby incorporated by reference in their entirety.

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).

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

US 11,080,758 B2

9

service the request (e.g., including a prompt for the information), or other information relating to an action based on the request. Furthermore, 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).

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.).

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, Wash. (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 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.

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

10

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.

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 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.

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.

It will be apparent that operation 320 may use various techniques to select advertisements based on voice-based

US 11,080,758 B2

11

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. (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.

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).

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 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.

12

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).

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.

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, 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.

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.

What is claimed is:

1. A method for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon, the method being implemented by one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method, the method comprising:

US 11,080,758 B2

13

providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;

receiving, by the one or more processors, words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine;

determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;

selecting, by the one or more processors, a purchase opportunity based on the determined context;

delivering, by the one or more processors, the selected purchase opportunity via an electronic device in communication with the one or more processors;

tracking, by the one or more processors, an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity, wherein the subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity;

building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and

interpreting, by the one or more processors, a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.

2. The method of claim 1, wherein the purchase opportunity relates to an item, the method further comprising:

receiving, by the one or more processors, a second natural language utterance;

determining, by the one or more processors, that the second natural language utterance comprises a request to purchase the item; and

completing, by the one or more processors, a purchase transaction of the item based on the second natural language utterance.

3. The method of claim 1, the method further comprising:

receiving, by the one or more processors, an indication of an occurrence of a user interaction with the purchase opportunity, wherein the interaction pattern includes the user interaction;

determining, by the one or more processors, whether the interpretation of the natural language utterance was correct based on the indication; and

interpreting, by the one or more processors, a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct.

4. The method of claim 1, wherein delivering the selected purchase opportunity via the electronic device comprises:

audibly presenting, by the electronic device, a natural language conversational purchase opportunity.

5. The method of claim 1, wherein the computer program instructions comprise at least a conversational language processor configured to interpret a natural language utterance, the method further comprising:

providing, by the one or more processors, the words or phrases as an input to the conversational language processor; and

receiving, by the one or more processors, an interpretation of the natural language utterance, based on the words or

14

phrases, as an output of the conversational language processor, wherein the purchase opportunity is selected based further on the interpretation of the natural language utterance.

6. The method of claim 5, wherein the conversational language processor comprises one or more domain agents, where a domain agent is configured to assist in:

(i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.

7. The method of claim 6, wherein the domains in which the one or more domain agents are configured to assist comprises an electronic commerce domain.

8. The method of claim 1, wherein recognizing the words or phrases from the natural language utterance comprises:

using, by the speech recognition engine, phonetic dictation to recognize a stream of phonemes in the natural language utterance, wherein the recognized words or phrases are based on the recognized stream of phonemes.

9. The method of claim 1, the method further comprising:

using, by the one or more processors, an environmental model to determine environmental information, wherein the context for the natural language utterance is determined based further on the environmental information.

10. The method of claim 1, wherein the environmental information comprises a user location, a user activity, or a user action.

11. The method of claim 1, further comprising:

invoking, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;

wherein the one or more system resources include at least an agent comprising an autonomous executable, wherein the executable comprises a redistributable package of functionality for a specific application that receives, processes, and responds to user questions, queries, and/or commands.

12. The method of claim 11, wherein the agent returns results of user questions, queries, and/or commands as responses to the user; and the responses are created using the results of information queries, system personality, and user preferences or other data in the user-specific profile.

13. The method of claim 11, further comprising updating the agent from a remote location over a network to provide new functionality.

14. The method of claim 11, wherein the agent uses system resources and the services of other agents.

15. The method of claim 11, further comprising providing a framework allowing the user to create or modify the agent.

16. The method of claim 14, further comprising providing a license management framework to enable the user to purchase the agent on a one time or subscription basis.

17. The method of claim 11, wherein:

the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories;

the user can interact with, and receive the same services and applications from, any one of the devices on the network; and

the location of a particular device addressed by the user is used as part of context for questions asked of the user.

18. The method of claim 1, further comprising:

capturing, by the one or more processors, the first natural language utterance;

US 11,080,758 B2

15

parsing, by the one or more processors, the first natural language utterance;
 interpreting, by the one or more processors, the first natural language utterance; and
 invoking, by the one or more processors, an applicable agent from a library of agents.

19. The method of claim 1, further comprising:
 invoking an agent to process the natural language utterance; and

creating a query to a local database or through a network interface to a remote database.

20. The method of claim 1, further comprising providing a user interface allowing the user to modify the user-specific profile, questions, queries, data sources, preferences, or parameters.

21. The method of claim 1, wherein the context comprises a categorization of the recognized words or phrases.

22. The method of claim 1, wherein the context comprises one of commands, directories, e-commerce, education, entertainment, lists, locations, multimedia, navigation, news, points of interest, search, stocks, recipes, scheduling, sports, subscriptions, transportation, or weather.

23. A system for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon, the system comprising:

one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to:

provide a natural language utterance as an input to a speech recognition engine;

receive words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine;

determine a context for the natural language utterance based on the recognized words or phrases;

select a purchase opportunity based on the determined context;

deliver the selected purchase opportunity via an electronic device in communication with the one or more processors;

track an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity, wherein the subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity;

build or update a user-specific profile based on the tracked interaction pattern; and

interpret a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.

24. The system of claim 23, wherein the purchase opportunity relates to an item, and wherein the one or more processors are further programmed to:

receive a second natural language utterance;

determine that the second natural language utterance comprises a request to purchase the item; and
 complete a purchase transaction of the item based on the second natural language utterance.

25. The system of claim 23, wherein the one or more processors are further programmed to:

16

receive an indication of an occurrence of a user interaction with the purchase opportunity, wherein the interaction pattern includes the user interaction;

determine whether the interpretation of the natural language utterance was correct based on the indication; and

interpret a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct.

26. The system of claim 23, wherein to deliver the selected purchase opportunity, the electronic device is configured to:

audibly present a natural language conversational purchase opportunity.

27. The system of claim 23, wherein the computer program instructions comprise at least a conversational language processor configured to interpret a natural language utterance, wherein the one or more processors are further programmed to:

provide the words or phrases as an input to the conversational language processor; and

receive an interpretation of the natural language utterance, based on the words or phrases, as an output of the conversational language processor, wherein the purchase opportunity is selected based further on the interpretation of the natural language utterance.

28. The system of claim 27, wherein the conversational language processor comprises one or more domain agents, where a domain agent is configured to assist in:

(i) interpreting requests related to its domain; and (ii)
 determining a response to the requests related to its domain.

29. The system of claim 28, wherein the domains in which the one or more domain agents are configured to assist comprises an electronic commerce domain.

30. The system of claim 23, wherein to recognize the words or phrases from the natural language utterance, the speech recognition engine is configured to:

use phonetic dictation to recognize a stream of phonemes in the natural language utterance, wherein the recognized words or phrases are based on the recognized stream of phonemes.

31. The system of claim 23, wherein the one or more processors are further programmed to:

use an environmental model to determine environmental information, wherein the context for the natural language utterance is determined based further on the environmental information.

32. The system of claim 23, wherein the environmental information comprises a user location, a user activity, or a user action.

33. The system of claim 23, wherein the one or more processors are further programmed to:

invoke, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;

wherein the one or more system resources include at least an agent comprising an autonomous executable, wherein the executable comprises a redistributable package of functionality for a specific application that receives, processes, and responds to user questions, queries, and/or commands.

34. The system of claim 33, wherein the agent returns results of user questions, queries, and/or commands as responses to the user; and the responses are created using the results of information queries, system personality, and user preferences or other data in the user-specific profile.

US 11,080,758 B2

17

35. The system of claim 33, wherein the one or more processors are further programmed to update the agent from a remote location over a network to provide new functionality.

36. The system of claim 33, wherein the agent uses system resources and the services of other agents.

37. The system of claim 33, wherein the one or more processors are further programmed to provide a framework allowing the user to create or modify the agent.

38. The system of claim 36, wherein the one or more processors are further programmed to provide a license management framework to enable the user to purchase the agent on a one time or subscription basis.

39. The system of claim 33, wherein:
the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories;
the user can interact with, and receive the same services and applications from, any one of the devices on the network; and
the location of a particular device addressed by the user is used as part of context for questions asked of the user.

40. The system of claim 23, wherein the one or more processors are further programmed to:
capture, by the one or more processors, the first natural language utterance;

18

parse, by the one or more processors, the first natural language utterance;
interpret, by the one or more processors, the first natural language utterance; and
invoke, by the one or more processors, an applicable agent from a library of agents.

41. The system of claim 23, wherein the one or more processors are further programmed to:
invoke an agent to process the natural language utterance;
and
create a query to a local database or through a network interface to a remote database.

42. The system of claim 23, wherein the one or more processors are further programmed to provide a user interface allowing the user to modify the user-specific profile, questions, queries, data sources, preferences, or parameters.

43. The system of claim 23, wherein the context comprises a categorization of the recognized words or phrases.

44. The system of claim 23, wherein the context comprises one of commands, directories, e-commerce, education, entertainment, lists, locations, multimedia, navigation, news, points of interest, search, stocks, recipes, scheduling, sports, subscriptions, transportation, or weather.

* * * * *

EXHIBIT E

(12) **United States Patent**
Kennewick et al.

(10) **Patent No.:** **US 9,502,025 B2**
(45) **Date of Patent:** **Nov. 22, 2016**

(54) **SYSTEM AND METHOD FOR PROVIDING A
NATURAL LANGUAGE CONTENT
DEDICATION SERVICE**

(75) Inventors: **Mike Kennewick**, Bellevue, WA (US);
Lynn Elise Armstrong, Woodinville,
WA (US)

(73) Assignee: **VoiceBox Technologies Corporation**,
Bellevue, WA (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 1345 days.

(21) Appl. No.: **12/943,699**

(22) Filed: **Nov. 10, 2010**

(65) **Prior Publication Data**

US 2011/0112921 A1 May 12, 2011

Related U.S. Application Data

(60) Provisional application No. 61/259,820, filed on Nov.
10, 2009.

(51) **Int. Cl.**
G10L 15/00 (2013.01)
G10L 15/18 (2013.01)
G06Q 30/06 (2012.01)
G06F 17/28 (2006.01)

(52) **U.S. Cl.**
CPC **G10L 15/18** (2013.01); **G06F 17/28**
(2013.01); **G06Q 30/0601** (2013.01)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,430,669 A 2/1984 Cheung 358/122
5,027,406 A 6/1991 Roberts et al. 381/43

5,155,743 A 10/1992 Jacobs 375/28
5,208,748 A 5/1993 Flores et al. 364/419
5,274,560 A 12/1993 LaRue 364/444
5,357,596 A 10/1994 Takebayashi et al. 395/2.84
5,377,350 A 12/1994 Skinner 395/600
5,386,556 A 1/1995 Hedin et al. 395/600
5,424,947 A 6/1995 Nagao et al. 364/419.08
5,471,318 A 11/1995 Ahuja et al. 358/400

(Continued)

FOREIGN PATENT DOCUMENTS

EP 1 320 043 A2 6/2003
EP 1 646 037 4/2006

(Continued)

OTHER PUBLICATIONS

Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles
Schwab & Co., Inc., Jul. 28, 2002, 1 page.

(Continued)

Primary Examiner — John Hayes

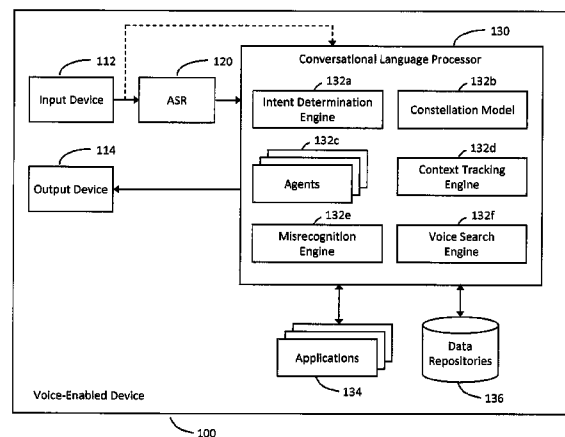
Assistant Examiner — Tim Hale

(74) *Attorney, Agent, or Firm* — Sheppard Mullin Richter
& Hampton LLP

(57) **ABSTRACT**

The system and method described herein may provide a
natural language content dedication service in a voice ser-
vices environment. In particular, providing the natural lan-
guage content dedication service may generally include
detecting multi-modal device interactions that include
requests to dedicate content, identifying the content
requested for dedication from natural language utterances
included in the multi-modal device interactions, processing
transactions for the content requested for dedication, pro-
cessing natural language to customize the content for recipi-
ents of the dedications, and delivering the customized con-
tent to the recipients of the dedications.

19 Claims, 8 Drawing Sheets



US 9,502,025 B2

Page 2

(56)

References Cited

U.S. PATENT DOCUMENTS

5,475,733 A	12/1995	Eisdorfer et al.	379/52	6,101,241 A	8/2000	Boyce et al.	379/88.01
5,488,652 A	1/1996	Bielby et al.	379/88	6,108,631 A	8/2000	Ruhl	704/270
5,499,289 A	3/1996	Bruno et al.	379/220	6,119,087 A	9/2000	Kuhn et al.	704/270
5,500,920 A	3/1996	Kupiec	395/2.79	6,122,613 A	9/2000	Baker	704/235
5,517,560 A	5/1996	Greenspan	379/114	6,134,235 A	10/2000	Goldman et al.	370/352
5,533,108 A	7/1996	Harris et al.	379/201	6,144,667 A	11/2000	Doshi et al.	370/401
5,537,436 A	7/1996	Bottoms et al.	375/222	6,144,938 A	11/2000	Surace et al.	704/257
5,539,744 A	7/1996	Chu et al.	370/60	6,154,526 A	11/2000	Dahlke et al.	379/88.03
5,557,667 A	9/1996	Bruno et al.	379/201	6,160,883 A	12/2000	Jackson et al.	379/230
5,563,937 A	10/1996	Bruno et al.	379/201	6,167,377 A	12/2000	Gillick et al.	704/240
5,577,165 A	11/1996	Takebayashi et al.	395/2.84	6,173,266 B1	1/2001	Marx et al.	704/270
5,590,039 A	12/1996	Ikeda et al.	395/759	6,173,279 B1	1/2001	Levin et al.	707/5
5,608,635 A	3/1997	Tamai	364/449.3	6,175,858 B1	1/2001	Bulfer et al.	709/206
5,617,407 A	4/1997	Bareis	369/275.3	6,185,535 B1	2/2001	Hedin et al.	704/270
5,633,922 A	5/1997	August et al.	379/220	6,188,982 B1	2/2001	Chiang	704/256
5,675,629 A	10/1997	Raffel et al.	379/58	6,192,110 B1	2/2001	Abella et al.	379/88.01
5,696,965 A	12/1997	Dedrick	395/610	6,192,338 B1	2/2001	Haszto et al.	704/257
5,708,422 A	1/1998	Blonder et al.	340/825.34	6,195,651 B1	2/2001	Dudemaine et al.	704/231
5,721,938 A	2/1998	Stuckey	395/754	6,199,043 B1	2/2001	Handel et al.	707/2
5,722,084 A	2/1998	Chakrin et al.	455/551	6,208,964 B1	3/2001	Happ	704/272
5,740,256 A	4/1998	Castello Da Costa et al.	361/94.7	6,208,972 B1	3/2001	Sabourin	704/244
5,742,763 A	4/1998	Jones	395/200.3	6,219,346 B1	3/2001	Grant et al.	704/275
5,748,841 A	5/1998	Morin et al.	395/2.66	6,219,643 B1	4/2001	Maxemchuk	370/338
5,748,974 A	5/1998	Johnson	395/759	6,226,612 B1	4/2001	Cohen et al.	704/257
5,752,052 A	5/1998	Richardson et al.	395/759	6,233,556 B1	5/2001	Srenger et al.	704/256
5,754,784 A	5/1998	Garland et al.	395/200.49	6,233,559 B1	5/2001	Teunen et al.	704/250
5,761,631 A	6/1998	Nasukawa	704/9	6,233,561 B1	5/2001	Balakrishnan	704/275
5,774,859 A	6/1998	Houser et al.	704/275	6,236,968 B1	5/2001	Junqua et al.	704/277
5,794,050 A	8/1998	Dahlgren et al.	395/708	6,246,981 B1	5/2001	Kanevsky et al.	704/275
5,794,196 A	8/1998	Yegnanarayanan et al.	704/255	6,246,990 B1	6/2001	Papineni et al.	704/235
5,797,112 A	8/1998	Komatsu et al.	701/201	6,266,636 B1	6/2001	Happ	704/275
5,799,276 A	8/1998	Komissarchik et al.	704/251	6,269,336 B1	7/2001	Kosaka et al.	704/244
5,802,510 A	9/1998	Jones	707/2	6,272,455 B1	7/2001	Ladd et al.	704/270
5,832,221 A	11/1998	Jones	375/200.36	6,272,455 B1	8/2001	Hoshen et al.	704/1
5,839,107 A	11/1998	Gupta et al.	704/270	6,275,231 B1	8/2001	Obradovich	345/349
5,848,396 A	12/1998	Gerace	705/10	6,278,968 B1	8/2001	Franz et al.	704/3
5,855,000 A	12/1998	Waibel et al.	704/235	6,288,319 B1	9/2001	Catona	84/609
5,867,817 A	2/1999	Catallo et al.	704/255	6,292,767 B1	9/2001	Jackson et al.	704/1
5,878,385 A	3/1999	Bralich et al.	704/9	6,301,560 B1	10/2001	Masters	704/251
5,878,386 A	3/1999	Coughlin	704/10	6,308,151 B1	10/2001	Smith	704/235
5,892,813 A	4/1999	Morin et al.	379/88.01	6,314,402 B1	11/2001	Monaco et al.	704/275
5,892,900 A	4/1999	Ginter et al.	726/26	6,356,869 B1	3/2002	Chapados et al.	704/275
5,895,464 A	4/1999	Bhandari et al.	707/3	6,362,748 B1	3/2002	Huang	340/901
5,895,466 A	4/1999	Goldberg et al.	707/5	6,366,882 B1	4/2002	Bijl et al.	704/235
5,897,613 A	4/1999	Chan	704/210	6,366,886 B1	4/2002	Dragosh et al.	704/270.1
5,902,347 A	5/1999	Backman et al.	701/200	6,374,214 B1	4/2002	Friedland et al.	704/235
5,911,120 A	6/1999	Jarett et al.	455/417	6,377,913 B1	4/2002	Coffman et al.	704/8
5,918,222 A	6/1999	Fukui et al.	707/1	6,381,535 B1	4/2002	Durocher et al.	701/202
5,926,784 A	7/1999	Richardson et al.	704/9	6,385,596 B1	5/2002	Wiser et al.	705/51
5,933,822 A	8/1999	Braden-Harder et al.	707/5	6,385,646 B1	5/2002	Brown et al.	709/217
5,953,393 A	9/1999	Culbreth et al.	379/88.25	6,393,403 B1	5/2002	Majaniemi	704/275
5,960,397 A	9/1999	Rahim	704/244	6,393,428 B1	5/2002	Miller et al.	707/102
5,960,399 A	9/1999	Barclay et al.	704/270	6,397,181 B1	5/2002	Li et al.	704/256
5,960,447 A	9/1999	Holt et al.	707/500	6,404,878 B1	6/2002	Jackson et al.	379/221.01
5,963,894 A	10/1999	Richardson et al.	704/9	6,405,170 B1	6/2002	Phillips et al.	704/270
5,963,940 A	10/1999	Liddy et al.	707/5	6,408,272 B1	6/2002	White et al.	704/270.1
5,987,404 A	11/1999	Della Pietra et al.	704/9	6,411,810 B1	6/2002	Maxemchuk	455/453
5,991,721 A	11/1999	Asano et al.	704/257	6,415,257 B1	7/2002	Junqua et al.	704/275
5,995,119 A	11/1999	Cosatto et al.	345/473	6,418,210 B1	7/2002	Sayko	379/142.15
5,995,928 A	11/1999	Nguyen et al.	704/251	6,420,975 B1	7/2002	DeLine et al.	340/815.4
6,009,382 A	12/1999	Martino et al.	704/1	6,429,813 B2	8/2002	Feigen	342/357.13
6,014,559 A	1/2000	Amin	455/413	6,430,285 B1	8/2002	Bauer et al.	379/265.01
6,018,708 A	1/2000	Dahan et al.	704/244	6,430,531 B1	8/2002	Polish	704/257
6,021,384 A	2/2000	Gorin et al.	704/1	6,434,523 B1	8/2002	Monaco	704/257
6,035,267 A	3/2000	Watanabe et al.	704/1	6,434,524 B1	8/2002	Weber	704/257
6,044,347 A	3/2000	Abella et al.	704/272	6,434,529 B1	8/2002	Walker et al.	704/275
6,049,602 A	4/2000	Foladare et al.	379/265	6,442,522 B1	8/2002	Carberry et al.	704/257
6,049,607 A	4/2000	Marash et al.	379/410	6,446,114 B1	9/2002	Bulfer et al.	709/206
6,058,187 A	5/2000	Chen	380/21	6,453,153 B1	9/2002	Bowker et al.	455/67.4
6,067,513 A	5/2000	Ishimitsu	704/233	6,453,292 B2	9/2002	Ramaswamy et al.	704/235
6,076,059 A	6/2000	Glickman et al.	704/260	6,456,711 B1	9/2002	Cheung et al.	379/265.09
6,078,886 A	6/2000	Dragosh et al.	704/270	6,456,974 B1	9/2002	Baker et al.	704/270.1
6,081,774 A	6/2000	De Hita et al.	704/9	6,466,654 B1	10/2002	Cooper et al.	379/88.01
6,085,186 A	7/2000	Christianson et al.	707/3	6,466,899 B1	10/2002	Yano et al.	704/1
				6,470,315 B1	10/2002	Netsch et al.	704/256
				6,498,797 B1	12/2002	Anerousis et al.	370/522
				6,499,013 B1	12/2002	Weber	704/257
				6,501,833 B2	12/2002	Phillips et al.	379/88.07
				6,501,834 B1	12/2002	Milewski et al.	379/93.24

US 9,502,025 B2

Page 3

(56)

References Cited

U.S. PATENT DOCUMENTS

6,505,155 B1	1/2003	Vanbuskirk et al.	704/246	6,901,366 B1	5/2005	Kuhn et al.	704/275
6,510,417 B1	1/2003	Woods et al.	704/275	6,910,003 B1	6/2005	Arnold et al.	704/4
6,513,006 B2	1/2003	Howard et al.	704/257	6,912,498 B2	6/2005	Stevens et al.	704/235
6,522,746 B1	2/2003	Marchok et al.	379/406.03	6,934,756 B2	8/2005	Maes	709/227
6,523,061 B1	2/2003	Halverson et al.	709/202	6,937,977 B2	8/2005	Gerson	704/201
6,532,444 B1	3/2003	Weber	704/257	6,944,594 B2	9/2005	Busayapongchai et al. .	704/275
6,539,348 B1	3/2003	Bond et al.	704/9	6,950,821 B2	9/2005	Faybishenko et al.	707/10
6,549,629 B2	4/2003	Finn et al.	381/92	6,954,755 B2	10/2005	Reisman	707/10
6,553,372 B1	4/2003	Brassell et al.	707/5	6,959,276 B2	10/2005	Droppo et al.	704/226
6,556,970 B1	4/2003	Sasaki et al.	704/257	6,968,311 B2	11/2005	Knockeart et al.	704/270
6,556,973 B1	4/2003	Lewin	704/277	6,973,387 B2	12/2005	Masclet et al.	701/211
6,560,576 B1	5/2003	Cohen et al.	704/270	6,975,993 B1	12/2005	Keiller	704/275
6,560,590 B1	5/2003	Shwe et al.	706/55	6,980,092 B2	12/2005	Turnbull et al.	340/425.5
6,567,778 B1	5/2003	Chao Chang et al.	704/257	6,983,055 B2	1/2006	Luo	381/313
6,567,797 B1	5/2003	Schuetze et al.	707/2	6,990,513 B2	1/2006	Belfiore et al.	709/203
6,570,555 B1	5/2003	Prevost et al.	345/156	6,996,531 B2	2/2006	Korall et al.	704/270
6,570,964 B1	5/2003	Murveit et al.	379/67.1	7,003,463 B1	2/2006	Maes et al.	704/270.1
6,571,279 B1	5/2003	Herz et al.	709/217	7,016,849 B2	3/2006	Arnold et al.	704/275
6,574,597 B1	6/2003	Mohri et al.	704/251	7,020,609 B2	3/2006	Thriff et al.	704/270.1
6,574,624 B1	6/2003	Johnson et al.	707/5	7,024,364 B2	4/2006	Guerra et al.	704/270
6,578,022 B1	6/2003	Foulger et al.	706/45	7,027,586 B2	4/2006	Bushey et al.	379/265.09
6,581,103 B1	6/2003	Dengler	709/231	7,027,975 B1	4/2006	Pazandak et al.	704/9
6,584,439 B1	6/2003	Geilhufe et al.	704/270	7,035,415 B2	4/2006	Belt et al.	381/92
6,587,858 B1	7/2003	Strazza	707/102	7,043,425 B2	5/2006	Pao	704/211
6,591,239 B1	7/2003	McCall et al.	704/275	7,054,817 B2	5/2006	Shao	704/270
6,594,257 B1	7/2003	Doshi et al.	370/352	7,058,890 B2	6/2006	George et al.	715/728
6,594,367 B1	7/2003	Marash et al.	381/92	7,062,488 B1	6/2006	Reisman	707/8
6,598,018 B1	7/2003	Junqua	704/251	7,069,220 B2	6/2006	Coffman et al.	704/275
6,601,026 B2	7/2003	Appelt et al.	704/9	7,072,834 B2	7/2006	Zhou	704/244
6,604,075 B1	8/2003	Brown et al.	704/270.1	7,082,469 B2 *	7/2006	Gold et al.	709/231
6,604,077 B2	8/2003	Dragosh et al.	704/270.1	7,092,928 B1	8/2006	Elad et al.	706/60
6,606,598 B1	8/2003	Holthouse et al.	704/275	7,107,210 B2	9/2006	Deng et al.	704/226
6,611,692 B2	8/2003	Raffel et al.	455/552	7,107,218 B1	9/2006	Preston	704/270
6,614,773 B1	9/2003	Maxemchuk	370/337	7,110,951 B1	9/2006	Lemelson et al.	704/270
6,615,172 B1	9/2003	Bennett et al.	704/257	7,127,400 B2	10/2006	Koch	704/270.1
6,622,119 B1	9/2003	Ramaswamy et al.	704/9	7,130,390 B2	10/2006	Abburi	379/88.17
6,629,066 B1	9/2003	Jackson et al.	704/9	7,136,875 B2	11/2006	Anderson et al.	707/104.1
6,631,346 B1	10/2003	Karaorman et al.	704/9	7,137,126 B1	11/2006	Coffman et al.	719/328
6,631,351 B1	10/2003	Ramachandran et al.	704/270	7,143,037 B1	11/2006	Chestnut	704/251
6,633,846 B1	10/2003	Bennett et al.	704/257	7,143,039 B1	11/2006	Stifelman et al.	704/270
6,643,620 B1	11/2003	Contolini et al.	704/270	7,146,319 B2	12/2006	Hunt	704/254
6,650,747 B1	11/2003	Bala et al.	379/265.06	7,165,028 B2	1/2007	Gong	704/233
6,658,388 B1	12/2003	Kleindienst et al.	704/275	7,184,957 B2	2/2007	Brookes et al.	704/246
6,678,680 B1	1/2004	Woo	707/6	7,190,770 B2	3/2007	Ando et al.	379/88.01
6,681,206 B1	1/2004	Gorin et al.	704/243	7,197,069 B2	3/2007	Agazzi et al.	375/233
6,691,151 B1	2/2004	Cheyre et al.	709/202	7,197,460 B1	3/2007	Gupta et al.	704/270.1
6,701,294 B1	3/2004	Ball et al.	704/257	7,203,644 B2	4/2007	Anderson et al.	704/246
6,704,396 B2	3/2004	Parolkar et al.	379/88.17	7,206,418 B2	4/2007	Yang et al.	381/92
6,704,576 B1 *	3/2004	Brachman et al.	455/503	7,207,011 B2	4/2007	Mulvey et al.	715/812
6,704,708 B1	3/2004	Pickering	704/235	7,228,276 B2	6/2007	Omote et al.	704/243
6,708,150 B1	3/2004	Hirayama et al.	704/243	7,231,343 B1	6/2007	Treadgold et al.	704/9
6,721,001 B1	4/2004	Berstis	348/231.3	7,236,923 B1	6/2007	Gupta	704/9
6,721,706 B1	4/2004	Strubbe et al.	704/275	7,272,212 B2	9/2007	Eberle et al.	379/88.17
6,735,592 B1	5/2004	Neumann et al.	707/101	7,277,854 B2	10/2007	Bennett et al.	704/257
6,741,931 B1	5/2004	Kohut et al.	701/209	7,283,829 B2	10/2007	Christenson et al.	455/461
6,742,021 B1	5/2004	Halverson et al.	709/218	7,283,951 B2	10/2007	Marchisio et al.	704/9
6,745,161 B1	6/2004	Arnold et al.	704/7	7,289,606 B2	10/2007	Sibal et al.	379/52
6,751,591 B1	6/2004	Gorin et al.	704/257	7,299,186 B2	11/2007	Kuzunuki et al.	704/270.1
6,751,612 B1	6/2004	Schuetze et al.	707/4	7,301,093 B2	11/2007	Sater et al.	84/615
6,754,485 B1	6/2004	Obradovich et al.	455/414.1	7,305,381 B1	12/2007	Poppink et al.	
6,754,627 B2	6/2004	Woodward	704/235	7,321,850 B2	1/2008	Wakita	704/10
6,757,544 B2	6/2004	Rangarajan et al.	455/456.1	7,328,155 B2	2/2008	Endo et al.	704/251
6,757,718 B1	6/2004	Halverson et al.	709/218	7,337,116 B2	2/2008	Charlesworth et al.	704/254
6,795,808 B1	9/2004	Strubbe et al.	704/275	7,340,040 B1	3/2008	Saylor et al.	379/67.1
6,801,604 B2	10/2004	Maes et al.	379/88.17	7,366,285 B2	4/2008	Parolkar et al.	379/88.17
6,801,893 B1	10/2004	Backfried et al.	704/257	7,366,669 B2	4/2008	Nishitani et al.	704/256
6,829,603 B1	12/2004	Chai et al.	707/5	7,376,645 B2	5/2008	Bernard	707/3
6,832,230 B1	12/2004	Zilliagus et al.	707/203	7,386,443 B1	6/2008	Parthasarathy et al.	704/201
6,833,848 B1	12/2004	Wolff et al.	345/719	7,398,209 B2	7/2008	Kennewick et al.	704/255
6,850,603 B1	2/2005	Eberle et al.	379/88.16	7,406,421 B2	7/2008	Odinak et al.	704/275
6,856,990 B2	2/2005	Barile et al.	707/10	7,415,414 B2	8/2008	Azara et al.	704/270
6,865,481 B2	3/2005	Kawazoe et al.	701/211	7,421,393 B1	9/2008	Di Fabbri et al.	704/275
6,868,380 B2	3/2005	Kroeker	704/240	7,424,431 B2	9/2008	Greene et al.	704/270
6,877,001 B2	4/2005	Wolf et al.	707/3	7,447,635 B1	11/2008	Konopka et al.	704/275
6,877,134 B1	4/2005	Fuller et al.	715/500.1	7,454,608 B2	11/2008	Gopalakrishnan et al. .	713/100
				7,461,059 B2	12/2008	Richardson et al.	707/5
				7,472,020 B2	12/2008	Brulle-Drews	701/211
				7,472,060 B1	12/2008	Gorin et al.	704/240
				7,478,036 B2	1/2009	Shen et al.	704/9

US 9,502,025 B2

Page 4

(56)

References Cited

U.S. PATENT DOCUMENTS

7,487,088 B1	2/2009	Gorin et al.	704/240	8,983,839 B2	3/2015	Kennewick et al.	704/254
7,493,259 B2	2/2009	Jones et al.	704/257	9,015,049 B2	4/2015	Baldwin et al.	
7,493,559 B1	2/2009	Wolff et al.	715/727	2001/0039492 A1	11/2001	Nemoto	704/231
7,502,738 B2	3/2009	Kennewick et al.	704/257	2001/0041980 A1	11/2001	Howard et al.	704/270
7,516,076 B2	4/2009	Walker et al.	704/275	2001/0049601 A1	12/2001	Kroeker et al.	704/254
7,529,675 B2	5/2009	Maes	704/270.1	2001/0054087 A1	12/2001	Flom et al.	709/218
7,536,297 B2	5/2009	Byrd et al.	704/10	2002/0015500 A1	2/2002	Belt et al.	381/66
7,536,374 B2	5/2009	Au	706/55	2002/0022927 A1	2/2002	Lemelson et al.	701/301
7,542,894 B2	6/2009	Murata	704/9	2002/0029261 A1 *	3/2002	Shibata	709/219
7,546,382 B2	6/2009	Healey et al.	709/246	2002/0032752 A1 *	3/2002	Gold et al.	709/218
7,558,730 B2	7/2009	Davis et al.	704/235	2002/0035501 A1	3/2002	Handel et al.	705/10
7,574,362 B2	8/2009	Walker et al.	704/275	2002/0049805 A1	4/2002	Yamada et al.	709/202
7,577,244 B2	8/2009	Taschereau	379/218.01	2002/0065568 A1	5/2002	Silfvast et al.	700/94
7,606,708 B2	10/2009	Hwang	704/257	2002/0069059 A1	6/2002	Smith	704/257
7,620,549 B2	11/2009	Di Cristo et al.	704/257	2002/0069071 A1	6/2002	Knockeart et al.	704/275
7,634,409 B2	12/2009	Kennewick et al.	704/257	2002/0082911 A1	6/2002	Dunn et al.	705/14
7,640,006 B2	12/2009	Portman et al.	455/412.1	2002/0087326 A1	7/2002	Lee et al.	704/270.1
7,640,160 B2	12/2009	Di Cristo et al.	704/257	2002/0087525 A1	7/2002	Abbott et al.	707/3
7,640,272 B2 *	12/2009	Mahajan et al.		2002/0120609 A1	8/2002	Lang et al.	707/1
7,676,365 B2	3/2010	Hwang et al.	704/240	2002/0124050 A1	9/2002	Middeljans	709/203
7,676,369 B2	3/2010	Fujimoto et al.	704/270	2002/0133402 A1	9/2002	Faber et al.	705/14
7,684,977 B2	3/2010	Morikawa	704/211	2002/0138248 A1	9/2002	Corston-Oliver et al.	704/1
7,693,720 B2 *	4/2010	Kennewick et al.	704/275	2002/0143532 A1	10/2002	McLean et al.	704/235
7,729,916 B2	6/2010	Coffman et al.	704/270	2002/0143535 A1	10/2002	Kist et al.	704/251
7,729,918 B2	6/2010	Walker et al.	704/275	2002/0161646 A1	10/2002	Gailey et al.	705/14
7,729,920 B2	6/2010	Chaar et al.	704/275	2002/0173961 A1	11/2002	Guerra	704/258
7,748,021 B2	6/2010	Obradovich	725/105	2002/0184373 A1	12/2002	Maes	709/228
7,788,084 B2	8/2010	Brun et al.	704/7	2002/0188602 A1	12/2002	Stubler et al.	707/3
7,801,731 B2	9/2010	Odinak et al.	704/275	2002/0198714 A1	12/2002	Zhou	704/252
7,809,570 B2	10/2010	Kennewick et al.	704/257	2003/0014261 A1	1/2003	Kageyama	704/275
7,818,176 B2	10/2010	Freeman et al.		2003/0016835 A1	1/2003	Elko et al.	381/92
7,831,426 B2	11/2010	Bennett	704/252	2003/0046346 A1	3/2003	Mumick et al.	709/205
7,831,433 B1	11/2010	Belvin et al.	704/275	2003/0064709 A1	4/2003	Gailey et al.	455/412
7,873,519 B2	1/2011	Bennett	704/257	2003/0065427 A1	4/2003	Funk et al.	701/1
7,873,523 B2	1/2011	Potter et al.	704/275	2003/0088421 A1	5/2003	Maes et al.	704/270.1
7,881,936 B2	2/2011	Longe et al.	704/257	2003/0097249 A1	5/2003	Walker et al.	704/1
7,890,324 B2	2/2011	Bangalore et al.	704/231	2003/0110037 A1	6/2003	Walker et al.	704/257
7,894,849 B2	2/2011	Kass et al.	455/550.1	2003/0112267 A1	6/2003	Belrose	345/728
7,902,969 B2	3/2011	Obradovich	340/439	2003/0115062 A1	6/2003	Walker et al.	704/258
7,917,367 B2	3/2011	Di Cristo et al.	704/270.1	2003/0120493 A1	6/2003	Gupta	704/270.1
7,920,682 B2	4/2011	Byrne et al.	379/88.18	2003/0135488 A1	7/2003	Amir et al.	707/3
7,949,529 B2	5/2011	Weider et al.	704/270	2003/0144846 A1	7/2003	Denenberg et al.	704/277
7,949,537 B2	5/2011	Walker et al.	704/275	2003/0158731 A1	8/2003	Falcon et al.	704/231
7,953,732 B2	5/2011	Frank et al.	707/724	2003/0161448 A1	8/2003	Parolkar et al.	379/88.17
7,974,875 B1	7/2011	Quilici et al.	705/14.4	2003/0182132 A1	9/2003	Niemoeller	704/275
7,983,917 B2	7/2011	Kennewick et al.	704/257	2003/0204492 A1	10/2003	Wolf et al.	707/3
7,984,287 B2	7/2011	Gopalakrishnan et al.	713/100	2003/0206640 A1	11/2003	Malvar et al.	381/93
8,005,683 B2 *	8/2011	Tessel et al.	704/275	2003/0212550 A1	11/2003	Ubale	704/215
8,015,006 B2	9/2011	Kennewick et al.	704/236	2003/0212558 A1	11/2003	Matula	704/260
8,060,367 B2	11/2011	Keaveney	704/247	2003/0212562 A1	11/2003	Patel et al.	704/275
8,069,046 B2	11/2011	Kennewick et al.	704/257	2003/0225825 A1	12/2003	Healey et al.	709/203
8,073,681 B2	12/2011	Baldwin et al.	704/9	2003/0236664 A1	12/2003	Sharma	704/251
8,077,975 B2	12/2011	Ma et al.	382/187	2004/0006475 A1	1/2004	Ehlen et al.	704/270.1
8,082,153 B2	12/2011	Coffman et al.	704/270	2004/0025115 A1	2/2004	Stenel et al.	715/513
8,086,463 B2	12/2011	Ativanichayaphong et al.		2004/0044516 A1	3/2004	Kennewick et al.	704/5
8,112,275 B2	2/2012	Kennewick et al.	704/240	2004/0098245 A1	5/2004	Walker et al.	704/1
8,140,327 B2	3/2012	Kennewick et al.	704/226	2004/0117804 A1	6/2004	Scahill et al.	719/320
8,140,335 B2	3/2012	Kennewick et al.	704/257	2004/0140989 A1	7/2004	Papageorge	345/700
8,145,489 B2	3/2012	Freeman et al.	704/257	2004/0158555 A1 *	8/2004	Seedman et al.	707/3
8,150,694 B2	4/2012	Kennewick et al.	704/257	2004/0166832 A1	8/2004	Portman et al.	455/412.1
8,155,962 B2	4/2012	Kennewick et al.	704/257	2004/0167771 A1	8/2004	Duan et al.	704/10
8,170,867 B2	5/2012	Germain	704/9	2004/0193408 A1	9/2004	Hunt	704/209
8,195,468 B2	6/2012	Weider et al.	704/275	2004/0193420 A1 *	9/2004	Kennewick et al.	704/257
8,200,485 B1	6/2012	Lee	704/231	2004/0199375 A1	10/2004	Ehsani et al.	704/4
8,219,399 B2	7/2012	Lutz et al.	704/270	2004/0205671 A1	10/2004	Sukehiro et al.	715/532
8,219,599 B2	7/2012	Tunstall-Pedoe	707/955	2004/0243417 A9	12/2004	Pitts, III et al.	704/276
8,224,652 B2	7/2012	Wang et al.	704/275	2005/0015256 A1	1/2005	Kargman	704/272
8,255,224 B2	8/2012	Singleton et al.	704/275	2005/0021334 A1	1/2005	Iwahashi	704/240
8,326,627 B2	12/2012	Kennewick et al.	704/257	2005/0021470 A1	1/2005	Martin et al.	705/51
8,326,634 B2	12/2012	Di Cristo et al.	704/270.1	2005/0021826 A1	1/2005	Kumar	709/232
8,326,637 B2	12/2012	Baldwin et al.	704/275	2005/0033574 A1	2/2005	Kim et al.	704/251
8,332,224 B2	12/2012	Di Cristo et al.	704/257	2005/0043940 A1	2/2005	Elder	704/9
8,447,607 B2	5/2013	Weider et al.	704/250	2005/0080632 A1	4/2005	Endo et al.	704/277
8,452,598 B2	5/2013	Kennewick et al.	704/257	2005/0114116 A1	5/2005	Fiedler	704/201
				2005/0125232 A1	6/2005	Gadd	704/270.1
				2005/0137850 A1	6/2005	Odell	704/4
				2005/0137877 A1	6/2005	Oesterling et al.	704/275
				2005/0143994 A1	6/2005	Mori et al.	704/235
				2005/0216254 A1	9/2005	Gupta et al.	704/9

US 9,502,025 B2

Page 5

(56)

References Cited

U.S. PATENT DOCUMENTS

2005/0234727 A1	10/2005	Chiu	704/270.1
2005/0246174 A1	11/2005	DeGolia	704/270
2005/0283752 A1	12/2005	Fruchter et al.	717/100
2006/0041431 A1	2/2006	Maes	704/270.1
2006/0047509 A1	3/2006	Ding et al.	704/231
2006/0206310 A1	9/2006	Ravikumar et al.	704/9
2006/0217133 A1	9/2006	Christenson et al.	455/461
2006/0285662 A1	12/2006	Yin et al.	379/88.16
2007/0033005 A1	2/2007	Cristo et al.	704/9
2007/0033020 A1	2/2007	(Kelleher) Francois et al.	704/226
2007/0038436 A1	2/2007	Cristo et al.	704/9
2007/0043574 A1	2/2007	Coffman et al.	704/275
2007/0043868 A1	2/2007	Kumar et al.	709/226
2007/0050191 A1	3/2007	Weider et al.	704/275
2007/0055525 A1	3/2007	Kennewick et al.	704/257
2007/0073544 A1	3/2007	Millett et al.	704/277
2007/0078708 A1	4/2007	Yu et al.	705/14
2007/0078709 A1	4/2007	Rajaram	705/14
2007/0118357 A1	5/2007	Kasravi et al.	704/10
2007/0135101 A1	6/2007	Ramati et al.	455/414.1
2007/0146833 A1 *	6/2007	Satomi et al.	358/537
2007/0162296 A1	7/2007	Altberg et al.	705/1
2007/0179778 A1	8/2007	Gong et al.	704/9
2007/0186165 A1	8/2007	Maislos et al.	715/728
2007/0198267 A1	8/2007	Jones et al.	704/257
2007/0214182 A1	9/2007	Rosenberg	707/104.1
2007/0250901 A1	10/2007	McIntire et al.	725/146
2007/0265850 A1	11/2007	Kennewick et al.	704/257
2007/0299824 A1	12/2007	Pan et al.	707/3
2008/0034032 A1	2/2008	Healey et al.	709/203
2008/0065386 A1	3/2008	Cross et al.	704/270
2008/0091406 A1	4/2008	Baldwin et al.	704/4
2008/0103761 A1	5/2008	Printz et al.	704/9
2008/0109285 A1	5/2008	Reuther et al.	705/7
2008/0115163 A1	5/2008	Gilboa et al.	725/34
2008/0133215 A1	6/2008	Sarukkai	704/2
2008/0140385 A1 *	6/2008	Mahajan et al.	704/9
2008/0147396 A1	6/2008	Wang	704/246
2008/0147410 A1	6/2008	Odinak	704/270.1
2008/0162471 A1	7/2008	Bernard	707/5
2008/0177530 A1	7/2008	Cross et al.	704/4
2008/0189110 A1	8/2008	Freeman et al.	704/251
2008/0235023 A1	9/2008	Kennewick et al.	704/257
2008/0235027 A1	9/2008	Cross	704/270.1
2008/0319751 A1	12/2008	Kennewick et al.	704/257
2009/0052635 A1	2/2009	Jones et al.	379/88.02
2009/0067599 A1	3/2009	Agarwal et al.	379/114.21
2009/0076827 A1	3/2009	Bulitta et al.	704/275
2009/0106029 A1	4/2009	DeLine et al.	704/275
2009/0117885 A1	5/2009	Roth	455/414.3
2009/0144271 A1	6/2009	Richardson et al.	707/5
2009/0150156 A1	6/2009	Kennewick et al.	704/257
2009/0171664 A1	7/2009	Kennewick et al.	704/257
2009/0216540 A1 *	8/2009	Tessel et al.	704/275
2009/0271194 A1	10/2009	Davis et al.	704/235
2009/0273563 A1	11/2009	Pryor	345/157
2009/0276700 A1 *	11/2009	Anderson et al.	715/700
2009/0299745 A1	12/2009	Kennewick et al.	704/257
2009/0313026 A1	12/2009	Coffman et al.	704/275
2010/0023320 A1	1/2010	Di Cristo et al.	704/9
2010/0029261 A1	2/2010	Mikkelsen et al.	455/419
2010/0036967 A1 *	2/2010	Caine et al.	709/236
2010/0049501 A1	2/2010	Kennewick et al.	704/9
2010/0049514 A1	2/2010	Kennewick et al.	704/233
2010/0057443 A1	3/2010	Di Cristo et al.	704/9
2010/0063880 A1	3/2010	Atsmon et al.	705/14.53
2010/0094707 A1	4/2010	Freer	705/14.54
2010/0145700 A1	6/2010	Kennewick et al.	704/257
2010/0185512 A1	7/2010	Borger et al.	705/14.49
2010/0204986 A1	8/2010	Kennewick et al.	704/226
2010/0204994 A1	8/2010	Kennewick et al.	704/257
2010/0217604 A1	8/2010	Baldwin et al.	704/275
2010/0286985 A1	11/2010	Kennewick et al.	704/257
2010/0299142 A1	11/2010	Freeman et al.	704/9

2010/0312566 A1	12/2010	Odinak et al.	704/275
2011/0022393 A1	1/2011	Waller et al.	704/270
2011/0112827 A1	5/2011	Kennewick et al.	704/9
2011/0131036 A1	6/2011	Di Cristo et al.	704/9
2011/0131045 A1	6/2011	Cristo et al.	704/249
2011/0231182 A1	9/2011	Weider et al.	704/9
2011/0231188 A1	9/2011	Kennewick et al.	704/236
2012/0022857 A1	1/2012	Baldwin et al.	704/9
2012/0101809 A1	4/2012	Kennewick et al.	704/9
2012/0101810 A1	4/2012	Kennewick et al.	704/9
2012/0109753 A1	5/2012	Kennewick et al.	705/14.58
2012/0150636 A1	6/2012	Freeman et al.	705/14.49
2012/0278073 A1	11/2012	Weider et al.	704/235
2015/0066627 A1	3/2015	Freeman et al.	705/14.41
2015/0073910 A1	3/2015	Kennewick et al.	705/14.58
2015/0095159 A1	4/2015	Kennewick et al.	705/14.58

FOREIGN PATENT DOCUMENTS

WO	WO 99/46763	9/1999
WO	WO 00/21232	4/2000
WO	WO 00/46792	8/2000
WO	WO 01/78065	10/2001
WO	WO 2004/072954	8/2004
WO	WO 2007/019318	2/2007
WO	WO 2007/021587	2/2007
WO	WO 2007/027546	3/2007
WO	WO 2007/027989	3/2007
WO	WO 2008/098039	8/2008
WO	WO 2008/118195	10/2008
WO	WO 2009/075912	6/2009
WO	WO 2009/145796	12/2009
WO	WO 2010/096752	8/2010

OTHER PUBLICATIONS

Lin, Bor-shen, et al., "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History", ASRU'99, 1999, 4 pages.

Kuhn, Thomas, et al., "Hybrid In-Car Speech Recognition for Mobile Multimedia Applications", Vehicular Technology Conference, IEEE, Jul. 1999, pp. 2009-2013.

Belvin, Robert, et al., "Development of the HRL Route Navigation Dialogue System", Proceedings of the First International Conference on Human Language Technology Research, San Diego, 2001, pp. 1-5.

Lind, R., et al., "The Network Vehicle—A Glimpse into the Future of Mobile Multi-Media" *IEEE Aerosp. Electron. Systems Magazine*, vol. 14, No. 9, Sep. 1999, pp. 27-32.

Zhao, Yilin, "Telematics: Safe and Fun Driving", *IEEE Intelligent Systems*, vol. 17, Issue 1, 2002, pp. 10-14.

Chai et al., "MIND: A Semantics-Based Multimodal Interpretation Framework for Conversational System", *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, Jun. 2002, pp. 37-46.

Cheyner et al., "Multimodal Maps: An Agent-Based Approach", *International Conference on Cooperative Multimodal Communication (CMC/95)*, May 24-26, 1995, pp. 111-121.

Elio et al., "On Abstract Task Models and Conversation Policies" in Workshop on Specifying and Implementing Conversation Policies, *Autonomous Agents '99*, Seattle, 1999, 10 pages.

Turunen, "Adaptive Interaction Methods in Speech User Interfaces", Conference on Human Factors in Computing Systems, Seattle, Washington, 2001, pp. 91-92.

Mao, Mark Z., "Automatic Training Set Segmentation for Multi-Pass Speech Recognition", Department of Electrical Engineering, Stanford University, CA, copyright 2005, IEEE, pp. I-685 to I-688.

Vanhoucke, Vincent, "Confidence Scoring and Rejection Using Multi-Pass Speech Recognition", Nuance Communications, Menlo Park, CA, 2005, 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, 1998, 4 pages.

US 9,502,025 B2

Page 6

(56)

References Cited**OTHER PUBLICATIONS**

El Meliani et al., "A Syllabic-Filler-Based Continuous Speech Recognizer for Unlimited Vocabulary", Canadian Conference on Electrical and Computer Engineering, vol. 2, Sep. 5-8, 1995, pp. 1007-1010.

Arrington, Michael, "Google Redefines GPS Navigation Landscape: Google Maps Navigation for Android 2.0", TechCrunch, printed from the Internet <<http://www.techcrunch.com/2009/10/28/google-redefines-car-gps-navigation-google-maps-navigation-android/>>, Oct. 28, 2009, 4 pages.

Bazzi, Issam et al., "Heterogeneous Lexical Units for Automatic Speech Recognition: Preliminary Investigations", *Processing of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Jun. 5-9, 2000, XP010507574, pp. 1257-1260.

O'Shaughnessy, Douglas, "Interacting with Computers by Voice: Automatic Speech Recognition and Synthesis", *Proceedings of the IEEE*, vol. 91, No. 9, Sep. 1, 2003, XP011100665, pp. 1272-1305.

Statement in Accordance with the Notice from the European Patent Office dated Oct. 1, 2007 Concerning Business Methods (OJ EPO Nov. 2007, 592-593), XP002456252.

* cited by examiner

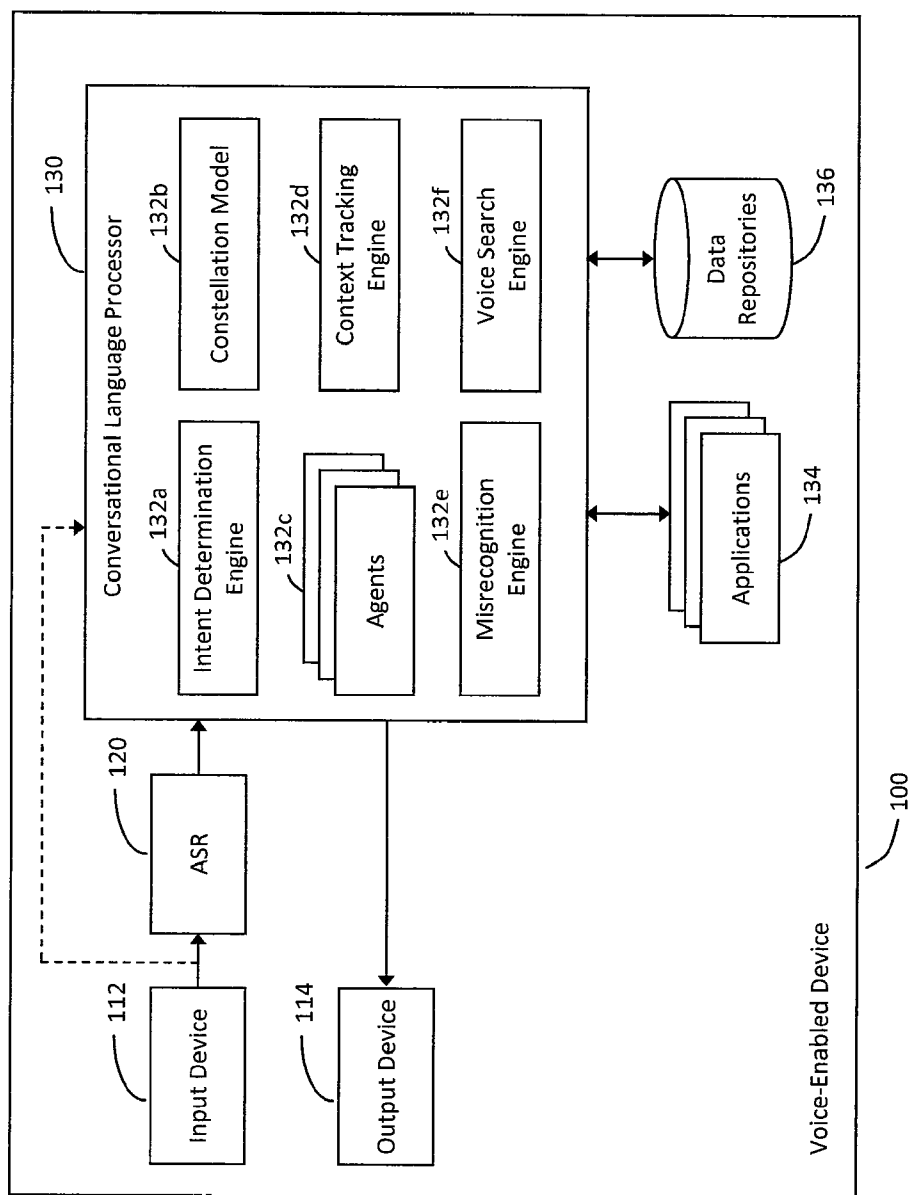


Figure 1

U.S. Patent

Nov. 22, 2016

Sheet 2 of 8

US 9,502,025 B2

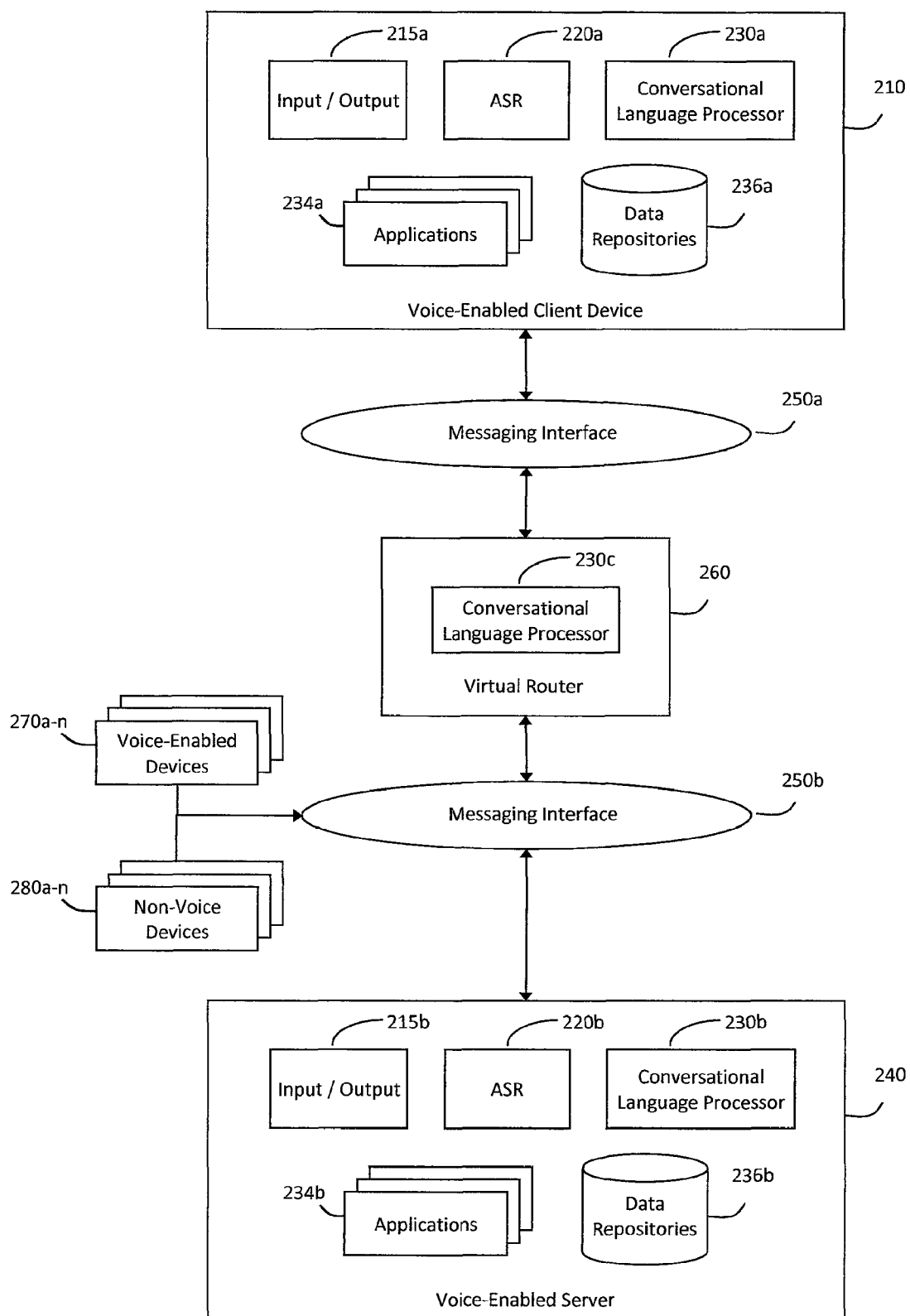


Figure 2

U.S. Patent

Nov. 22, 2016

Sheet 3 of 8

US 9,502,025 B2

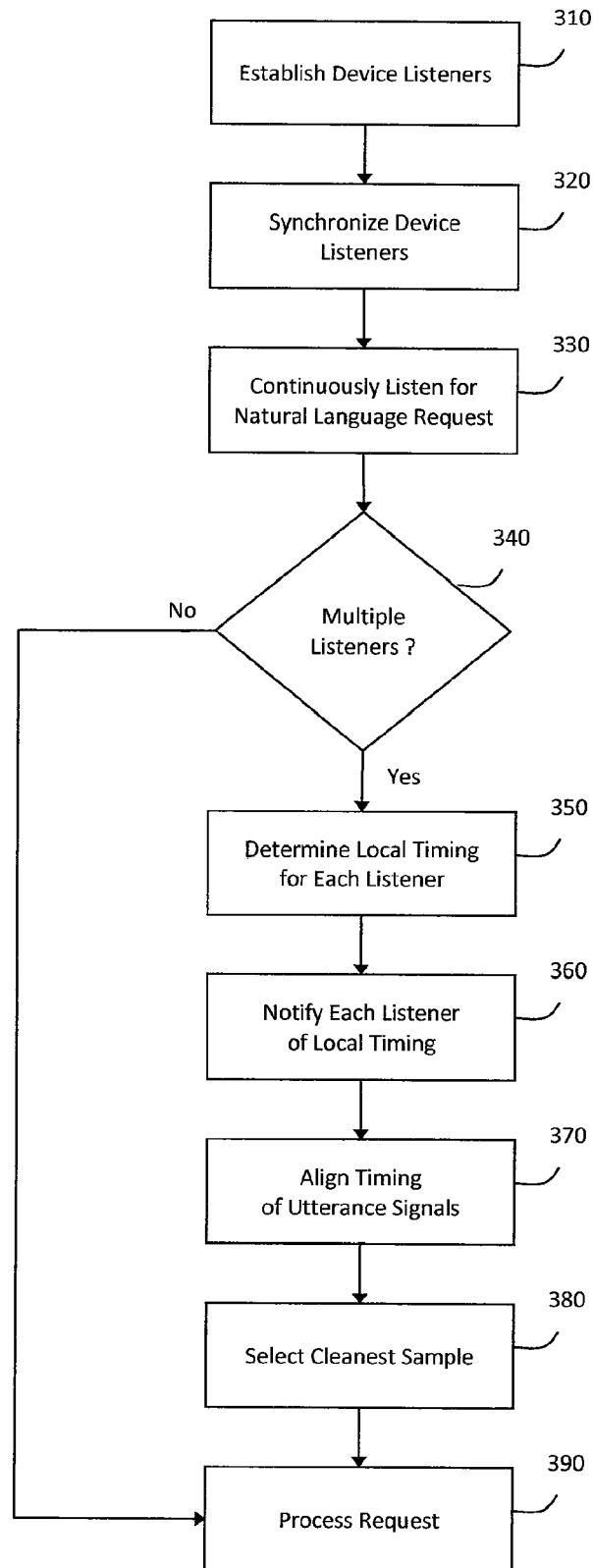
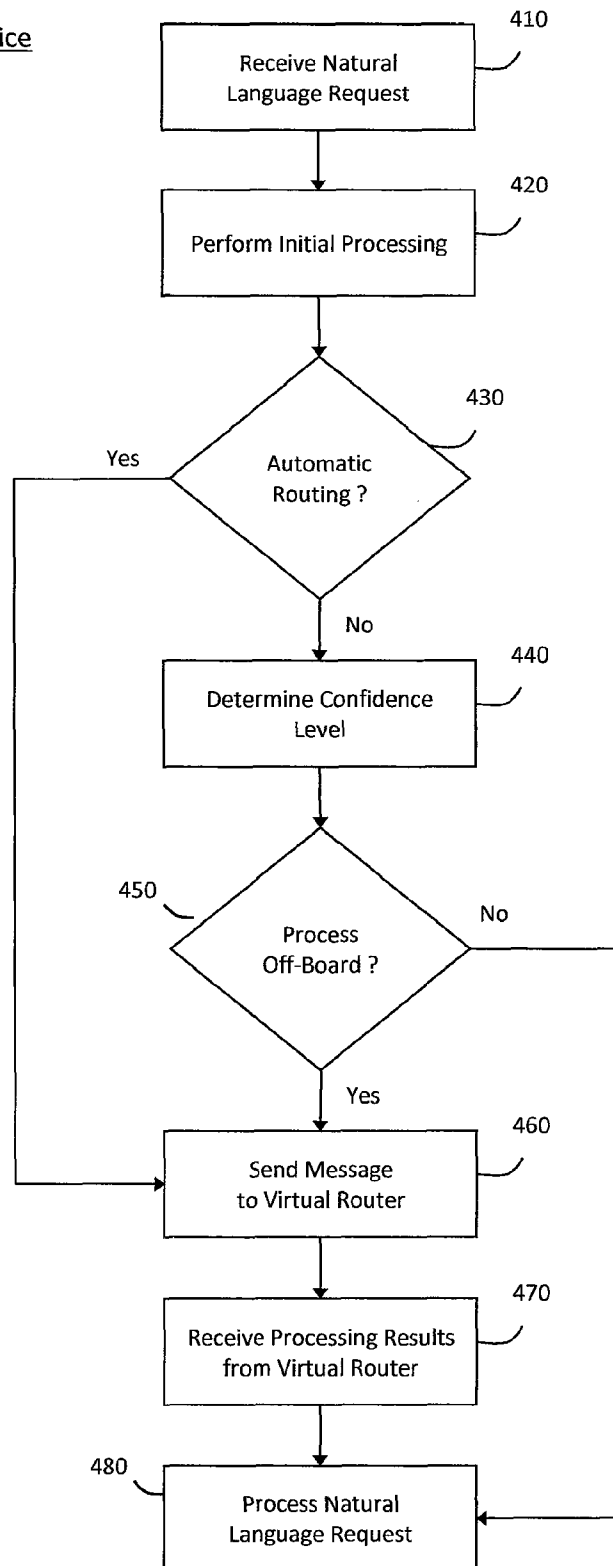
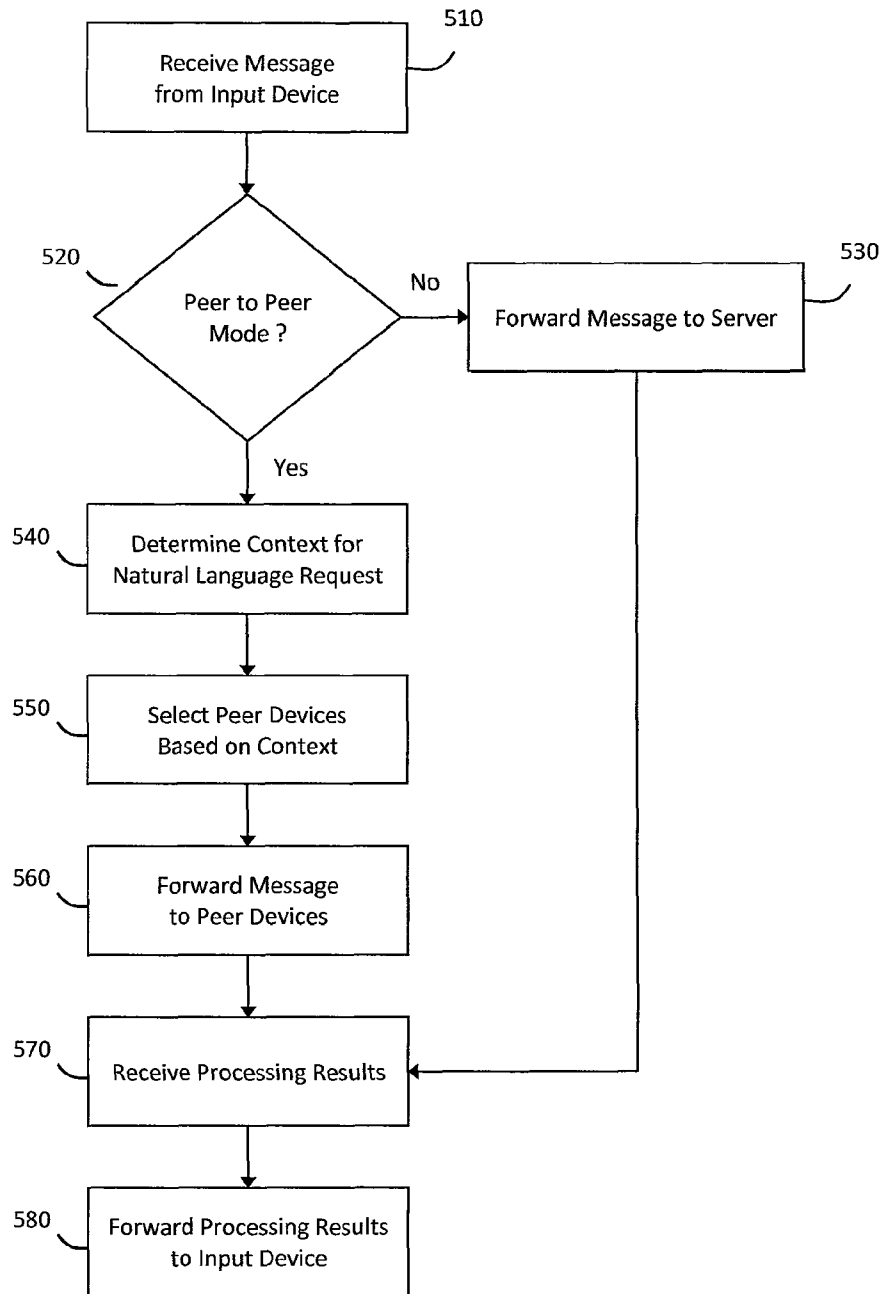


Figure 3

Client Device**Figure 4**

Virtual Router**Figure 5**

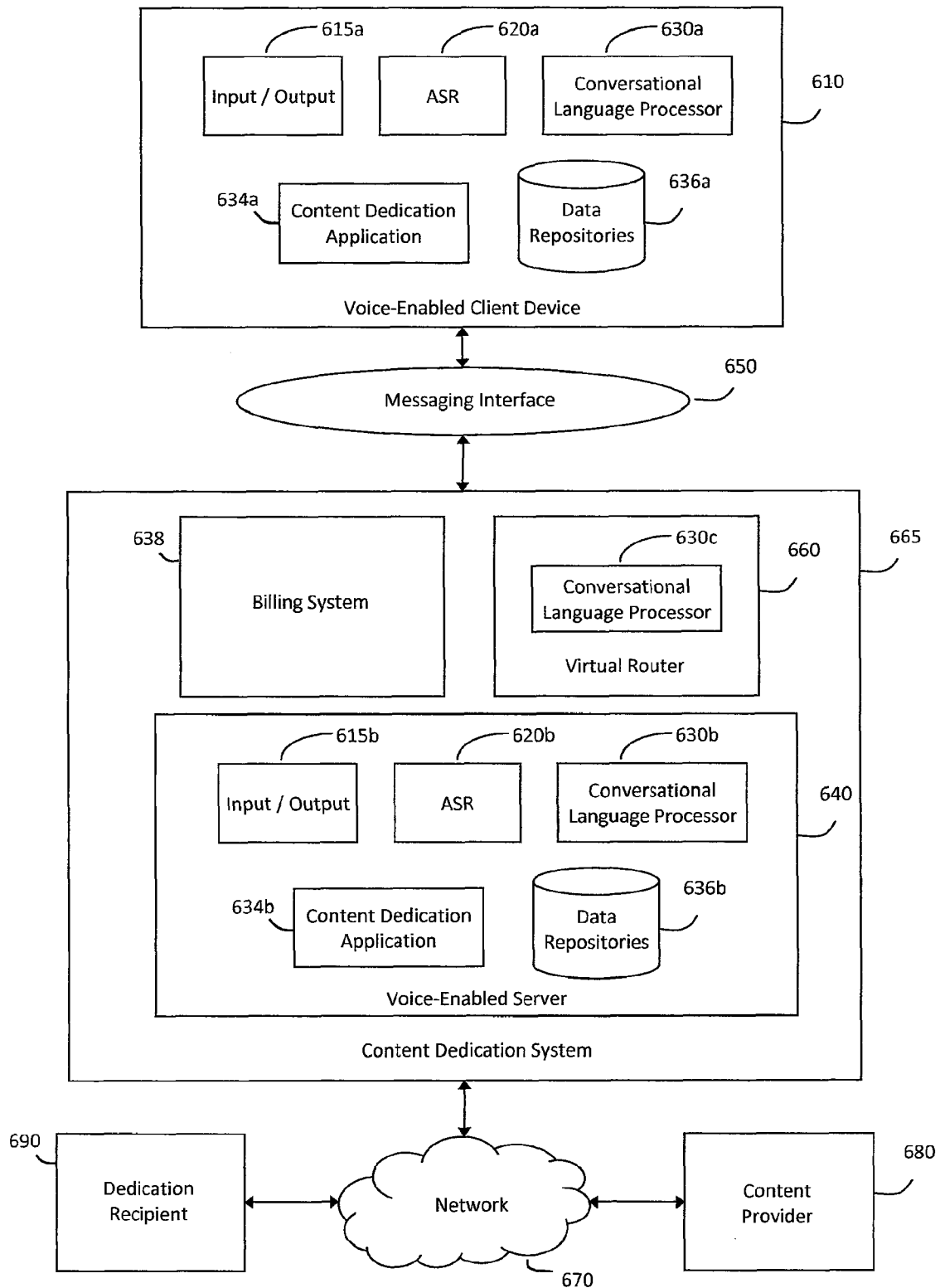


Figure 6

Client Device

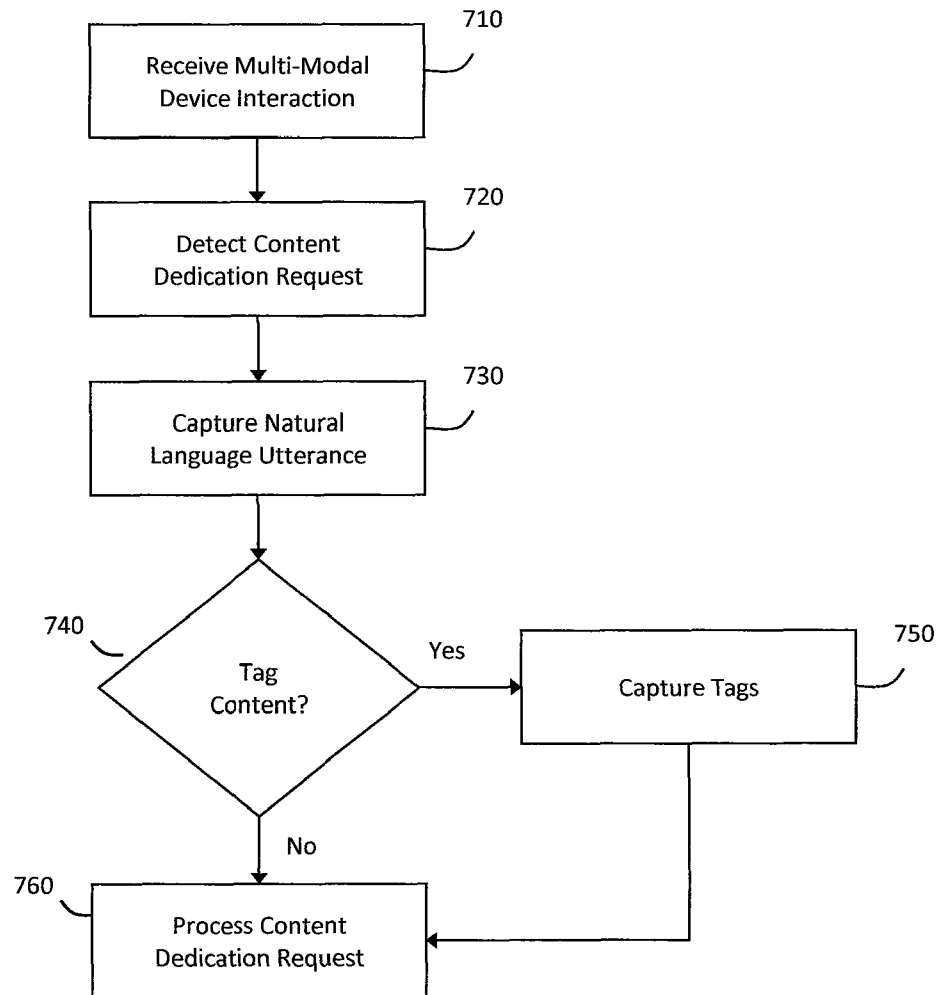


Figure 7

Content Dedication System

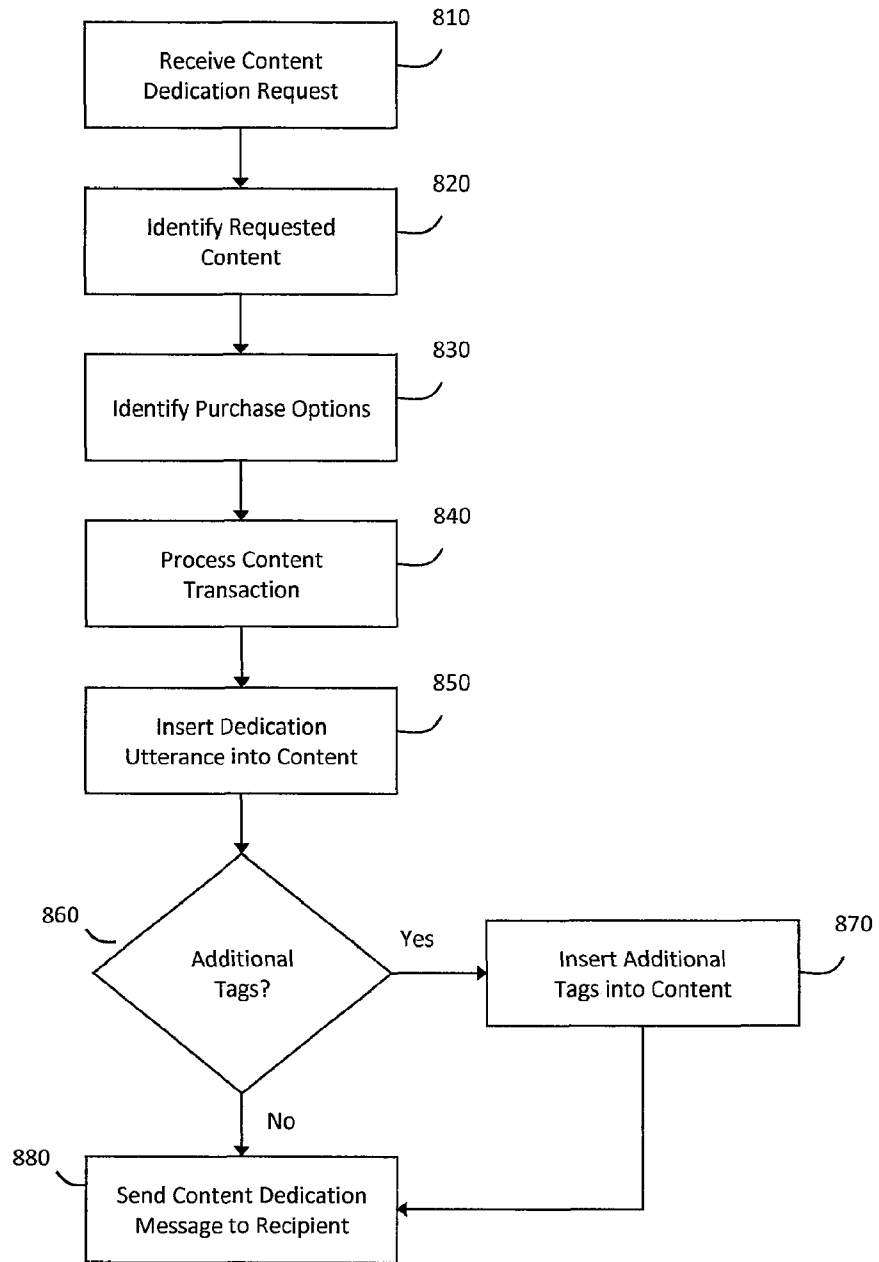


Figure 8

US 9,502,025 B2

1

SYSTEM AND METHOD FOR PROVIDING A NATURAL LANGUAGE CONTENT DEDICATION SERVICE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Ser. No. 61/259,820, entitled "System and Method for Providing a Natural Language Content Dedication Service," filed Nov. 10, 2009, the contents of which are hereby incorporated by reference in their entirety.

FIELD OF THE INVENTION

The invention generally relates to providing a natural language content dedication service in a voice services environment, and in particular, to detecting multi-modal device interactions that include requests to dedicate content, identifying the content requested for dedication from natural language utterances included in the multi-modal device interactions, processing transactions for the content requested for dedication, processing natural language to customize the content for recipients of the dedications, and delivering the customized content to the recipients of the dedications.

BACKGROUND OF THE INVENTION

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.

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 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.

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

2

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.

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.

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 a finite number of devices at any given time, yet content or services associated with users' devices other than those 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 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 cooperative multi-modal interactions with the device to simply utter words in natural language to request the function.

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. In another example, users often listen to music or interact with other media in mobile environments, such that interest in purchasing music, media, or other content may be fleeting or often occur on an impulse basis. Whereas existing human to machine interfaces that

US 9,502,025 B2

3

lack voice recognition capabilities typically fall short in providing mechanisms that can readily meet this demand, adding voice recognition to an electronic device can substantially simplify human to machine interaction in a manner that can meet user needs, improve experience, and satisfy potentially transient consumer interests. As such, interaction with electronic devices could be made far more efficient if users were provided with the ability to use natural language in order to exploit buried or otherwise difficult to use functionality.

SUMMARY OF THE INVENTION

According to one aspect of the invention, a system and method for providing a natural language content dedication service may generally operate in a voice services environment that includes one or more electronic devices that can receive multi-modal natural language device interactions. In particular, providing the natural language content dedication service may generally include detecting multi-modal device interactions that include requests to dedicate content, identifying the content requested for dedication from natural language utterances included in the multi-modal device interactions, processing transactions for the content requested for dedication, processing natural language to customize the content for recipients of the dedications, and delivering the customized content to the recipients of the dedications.

According to one aspect of the invention, the natural language content dedication service may operate in a hybrid processing environment, which may generally include a plurality of multi-modal devices configured to cooperatively interpret and process natural language utterances included in the multi-modal device interactions. For example, a virtual router may receive messages that include encoded audio corresponding to natural language utterances contained in the multi-modal device interactions, which may be received at one or more of the plurality of multi-modal devices in the hybrid processing environment. For example, the virtual router may analyze the encoded audio to select a cleanest sample of the natural language utterances and communicate with one or more other devices in the hybrid processing environment to determine an intent of the multi-modal device interactions. The virtual router may then coordinate resolving the multi-modal device interactions based on the intent of the multi-modal device interactions.

According to one aspect of the invention, a method for providing the natural language content dedication service may comprise detecting a multi-modal device interaction at an electronic device, wherein the multi-modal device interaction may include at least a natural language utterance. One or more messages containing information relating to the multi-modal device interaction may then be communicated to the virtual router through a messaging interface. The electronic device may then receive one or more messages (e.g., from the virtual router through message interface), wherein the messages may contain information relating to an intent of the multi-modal device interaction. As such, the multi-modal device interaction may be resolved at the electronic device based on the information contained in the one or more messages received from the virtual router.

According to one aspect of the invention, a system for providing the natural language content dedication service may generally include a voice-enabled client device that can communicate with a content dedication system through the messaging interface. The content dedication system include a voice-enabled server, which may be configured to com-

4

municate with the virtual router through another messaging interface, or the content dedication system may alternatively include the virtual router. In addition, the content dedication system may further include a billing system for processing transactions relating to the content dedication service. The natural language content dedication service may be provided on any suitable voice-enabled client device having a suitable combination of input and output devices that can receive and respond to multi-modal device interactions that include natural language utterances, and the input and output devices may be further arranged to receive and respond to any other suitable type of input and output.

According to one aspect of the invention, operating the natural language content dedication service may generally include a user of the voice-enabled client device listening to music, watching video, or otherwise interacting with content and providing a multi-modal natural language request to engage in a transaction to dedicate the music, video, or other content. Furthermore, the voice-enabled client device may be included within the hybrid processing environment that includes the plurality of multi-modal devices, whereby the content dedication request may relate to content played on a different device from the voice-enabled client device, although the content dedication request may relate to any suitable content (i.e., the request need not necessarily relate to played content, as users may provide natural language to request content dedications for any suitable content, including a particular song or video that the user may be thinking about).

According to one aspect of the invention, in response to the voice-enabled client device receiving a multi-modal interaction that includes a natural language utterance, the voice-enabled client device may invoke an Automatic Speech Recognizer (ASR) to generate a preliminary interpretation of the utterance. The ASR may then provide the preliminary interpretation of the utterance to a conversational language processor, which may attempt to determine an intent for the multi-modal interaction. For example, to determine the intent for the multi-modal interaction, the conversational language processor may determine a most likely context for the interaction from the preliminary interpretation of the utterance, any accompanying non-speech inputs in the multi-modal interaction that relate to the utterance, contexts associated with prior requests, short-term and long-term shared knowledge, or any other suitable information for interpreting the multi-modal interaction. Thus, in response to the conversational language processor determining that the intent for the multi-modal interaction relates to a content dedication request, a content dedication application may be invoked to resolve the content dedication request.

According to one aspect of the invention, to resolve the intent of the content dedication request, the conversational language processor may search one or more data repositories that contain content information to identify content matching criteria contained in the content dedication request. Moreover, the conversational language processor may further cooperate with other devices in the hybrid processing environment to search for or otherwise identify the content requested for dedication (e.g., in response to a local data repository not yielding adequate results, another device in the hybrid processing environment having a larger content data repository than the client device may be invoked.). The conversational language processor may then receive appropriate results identifying the content requested for dedication and present the results to the user through the output device (e.g., displaying information about the content, playing a

US 9,502,025 B2

5

sample clip of the content, displaying options to purchase the content, recommending similar content, etc.). The results presented through the output device may further include an option to confirm the content dedication, wherein the user may confirm the content dedication in a natural language utterance, a non-speech input, or any suitable combination thereof. The content dedication application may then be invoked to process the content dedication request.

According to one aspect of the invention, to process the content dedication request, the content dedication application may capture a natural language utterance that contains the dedication to accompany the content. The user may then provide the dedication utterance through the voice-enabled input device, and the dedication utterance may then be converted into an electronic signal that the content dedication application captures for the dedication. In addition, the content dedication application may prompt the user to provide any additional tags for the dedicated content (e.g., an image to insert as album art in the dedicated content, an utterance to insert or transcribe into metadata tags for the dedicated content, a non-speech or data input to insert in the metadata tags for the dedicated content, etc.). The content dedication application may further prompt the user to identify a recipient of the dedication, wherein the user may provide any suitable multi-modal input that includes information identifying the recipient of the content dedication. The content dedication application may then route the request to the content dedication system, which may process a transaction for the content dedication.

According to one aspect of the invention, processing the transaction for the content dedication may include the content dedication system receiving encoded audio corresponding to the dedication utterance through the messaging interface. The content dedication system may then insert the encoded audio corresponding to the dedication utterance within the dedicated content, verbally annotate the dedicated content with the encoded audio, and/or transcribe the dedication utterance into a textual annotation for the dedicated content. Similarly, any utterances to insert into the metadata tags for the dedicated content may provide further verbal annotations for the dedicated content, and any such utterances may also be transcribed into text to provide further textual annotations for the dedicated content. The content dedication system may then invoke a content dedication application hosted on the voice-enabled server to process the request. In particular, the content dedication application hosted on the voice-enabled server may identify the content requested for dedication (e.g., if the content dedication application on the voice-enabled client device was unable to suitably identify the requested content, or the information communicated from the voice-enabled client device may identify the requested content if the content dedication application on the voice-enabled client device was able to suitably identify the requested content, etc.). In response to identifying the content to be dedicated, the content dedication system may then communicate with the billing system to process an appropriate transaction for the dedication request based on a selected purchase option for the dedication request.

According to one aspect of the invention, the content dedication system may support various purchase options to provide users with flexibility in requesting content dedications. For example, a buy-to-own purchase option may include the content dedication system purchasing full rights to the content from an appropriate content provider. The billing system may then charge the user of the voice-enabled client device an appropriate amount that encompasses the

6

cost for purchasing the rights to the content from the content provider and a service charge for customizing the content with the dedication and any additional tags and subsequently delivering the customized content to the dedication recipient. The user may be charged in a similar manner under a pay-to-play purchase option, except that the rights purchased from the content provider may be limited (e.g., to a predetermined number of plays). Thus, the cost for purchasing the content from the content provider may be somewhat less under the pay-to-play purchase option, such that the billing system may charge the user somewhat less under the pay-to-play purchase option. Under a paid subscription purchase option, the user may pay a periodic service charge to the content dedication system that permits the user to make content dedications based on terms of the subscription (e.g., a predetermined number an unlimited number of content dedications may be made in a subscription period depending on the particular terms of the user's subscription). Furthermore, other purchase options may be suitably employed, as will be apparent.

According to one aspect of the invention, a service provider associated with the content dedication system may negotiate agreements with content providers to determine the manner in which revenues for content transactions will be shared between the content dedication system and the content providers. For example, an agreement may permit a particular content provider to keep all of the revenue for transactions that include purchasing content from the content provider, while the service provider associated with the content dedication system agrees to recoup any such costs from users. In another example, an agreement may share revenue for content transactions between a content provider and the service provider. Thus, the agreements may generally include any suitable arrangements that define the manner in which content providers and the service provider agree to divide revenue for transactions in, which the content dedication system purchases content from the content provider, while the service provider associated with the content dedication system may be responsible for billing users for any natural language aspects of the content dedication service (e.g., for inserting dedication utterances, transcribing utterances into metadata tags, delivering the content to recipients, etc.).

According to one aspect of the invention, in response to purchasing the requested content from an appropriate service provider and determining the appropriate billing options for the content dedication, the content dedication system may insert the natural language dedication utterance into the dedicated content, verbally annotate the dedicated content with the dedication utterance, or otherwise associate the content with the dedication utterance. Furthermore, the content dedication system may determine whether any additional tags have been specified for the dedicated content and insert such additional tags into the dedicated content, as appropriate (e.g., inserting an image or picture into metadata tags corresponding to album art for the dedicated content, transcribing natural language utterances, non-voice, and/or data inputs into text and inserting such text into the metadata tags for the dedicated content, etc.). The content dedication system may then send a content dedication message to the recipient of the dedication, wherein the message may include a link that the recipient can select to stream, download, or otherwise access the dedicated content, the dedication utterance, etc.). Thus, the content dedication message may generally notify the recipient that content has been dedicated to the recipient and provide various mechanisms for the recipient to access the content dedication, as will be

US 9,502,025 B2

7

apparent. Furthermore, if the purchase option selected for the content dedication includes the buy-to-own purchase option or another purchase option that confers full rights to the dedicated content, the recipient may then own the full rights to the dedicated content, otherwise the recipient may own rights with respect to the dedicated content based on whatever terms the selected purchase option provides.

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

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a block diagram of an exemplary voice-enabled device that can be used for hybrid processing in a natural language voice services environment, according to one aspect of the invention.

FIG. 2 illustrates a block diagram of an exemplary system for hybrid processing in a natural language voice service environment, according to one aspect of the invention.

FIG. 3 illustrates a flow diagram of an exemplary method for initializing various devices that cooperate to perform hybrid processing in a natural language voice services environment, according to one aspect of the invention.

FIGS. 4-5 illustrate flow diagrams of exemplary methods for hybrid processing in a natural language voice services environment, according to one aspect of the invention.

FIG. 6 illustrates a block diagram of an exemplary system for providing a natural language content dedication service, according to one aspect of the invention.

FIGS. 7-8 illustrate flow diagrams of exemplary methods for providing a natural language content dedication service, according to one aspect of the invention.

DETAILED DESCRIPTION

According to one aspect of the invention, FIG. 1 illustrates a block diagram of an exemplary voice-enabled device 100 that can be used for hybrid processing in a natural language voice services environment. As will be apparent from the further description to be provided herein, the voice-enabled device 100 illustrated in FIG. 1 may generally include an input device 112, or a combination of input devices 112, which may enable a user to interact with the voice-enabled device 100 in a multi-modal manner. In particular, the input devices 112 may generally include any suitable combination of at least one voice input device 112 (e.g., a microphone) and at least one non-voice input device 112 (e.g., a mouse, touch-screen display, wheel selector, etc.). As such, the input devices 112 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).

In one implementation, the voice-enabled device 100 may enable the user to engage in various multi-modal conversational interactions, which the voice-enabled device 100 may process in a free-form and cooperative manner to execute various tasks, resolve various queries, or otherwise resolve various natural language requests included in the multi-modal interactions. For example, in one implementation, the voice-enabled device 100 may include various natural language processing components, including at least a voice-click module coupled to the one or more input devices 112, as described in further detail in co-pending U.S. patent

8

application Ser. No. 12/389,678, entitled "System and Method for Processing Multi-Modal Device Interactions in a Natural Language Voice Services Environment," filed Feb. 20, 2009, the contents of which are hereby incorporated by reference in their entirety. Thus, as will be described in further detail herein, the one or more input devices 112 and the voice-click module may be collectively configured to process various multi-modal interactions between the user and the voice-enabled device 100.

For example, in one implementation, the multi-modal interactions may include at least one natural language utterance, wherein the natural language utterance may be converted into an electronic signal. The electronic signal may then be provided to an Automatic Speech Recognizer (ASR) 120, which may also be referred to as a speech recognition engine 120 and/or a multi-pass speech recognition engine 120. In response to receiving the electronic signal corresponding to the utterance, the ASR 120 may generate one or more preliminary interpretations of the utterance and provide the preliminary interpretation to a conversational language processor 130. Additionally, in one implementation, the multi-modal interactions may include one or more non-voice interactions with the one or more input devices 112 (e.g., button pushes, multi-touch gestures, point of focus or attention focus selections, etc.). As such, the voice-click module may extract context from the non-voice interactions and provide the context to the conversational language processor 130 for use in generating an interpretation of the utterance (i.e., via the dashed line illustrated in FIG. 1). As such, as described in greater detail below, the conversational language processor 130 may analyze the utterance and any accompanying non-voice interactions to determine an intent of the multi-modal interactions with the voice-enabled device 100.

In one implementation, as noted above, the voice-enabled device 100 may include various natural language processing components that can support free-form utterances and/or other forms of non-voice device interactions, which may liberate the user from restrictions relating to the manner of formulating commands, queries, or other requests. As such, the user may provide the utterance to the voice input device 112 using any manner of speaking, and may further provide other non-voice interactions to the non-voice input device 112 to request any content or service available through the voice-enabled device 100. For instance, in one implementation, in response to receiving the utterance at the voice input device 112, the utterance may be processed using techniques described in U.S. patent application Ser. No. 10/452,147, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," which issued as U.S. Pat. No. 7,398,209 on Jul. 8, 2008, and co-pending U.S. patent application Ser. No. 10/618,633, entitled "Mobile Systems and Methods for Responding to Natural Language Speech Utterance," filed Jun. 15, 2003, the contents 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 112 to provide buttons pushes, multi-touch gestures, point of focus or attention focus selections, or other non-voice device interactions, which may provide further context or other information relating to the natural language utterances and/or the requested content or service.

In one implementation, the voice-enabled device 100 may be coupled to one or more additional systems that may be configured to cooperate with the voice-enabled device 100 to interpret or otherwise process the multi-modal interactions that include combinations of natural language utter-

US 9,502,025 B2

9

ances and/or non-voice device interactions. For example, as will be described in greater detail below in connection with FIG. 2, the one or more additional systems may include one or more multi-modal voice-enabled devices having similar natural language processing capabilities to the voice-enabled device 100, one or more non-voice devices having data retrieval and/or task execution capabilities, and a virtual router that coordinates interaction among the voice-enabled device 100 and the additional systems. As such, the voice-enabled device 100 may include an interface to an integrated natural language voice services environment that includes a plurality of multi-modal devices, wherein the user may request content or services available through any of the multi-modal devices.

For example, in one implementation, the conversational language processor 130 may include a constellation model 132b that provides knowledge relating to content, services, applications, intent determination capabilities, and other features available in the voice services environment, as described in co-pending U.S. patent application Ser. No. 12/127,343, entitled "System and Method for an Integrated, Multi-Modal, Multi-Device Natural Language Voice Services Environment," filed May 27, 2008, the contents of which are hereby incorporated by reference in their entirety. As such, the voice-enabled device 100 may have access to shared knowledge relating to natural language processing capabilities, context, prior interactions, domain knowledge, short-term knowledge, long-term knowledge, and cognitive models for the various systems and multi-modal devices, providing a cooperative environment for resolving the multi-modal interactions received at the voice-enabled device 100.

In one implementation, the input devices 112 and the voice-click module coupled thereto may be configured to continually monitor for one or more multi-modal interactions received at the voice-enabled device 100. In particular, the input devices 112 and the voice-click module may continually monitor for one or more natural language utterances and/or one or more distinguishable non-voice device interactions, which may collectively provide the relevant context for retrieving content, executing tasks, invoking services or commands, or processing any other suitable requests. Thus, in response to detecting one or more multi-modal interactions, the input devices 112 and/or the voice-click module may signal the voice-enabled device 100 that an utterance and/or a non-voice interaction have been received. For example, in one implementation, the non-voice interaction may provide context for sharpening recognition, interpretation, and understanding of an accompanying utterance, and moreover, the utterance may provide further context for enhancing interpretation of the accompanying non-voice interaction. Accordingly, the utterance and the non-voice interaction may collectively provide relevant context that various natural language processing components may use to determine an intent of the multi-modal interaction that includes the utterance and the non-voice interaction.

In one implementation, as noted above, processing the utterance included in the multi-modal interaction may be initiated at the ASR 120, wherein the ASR 120 may generate one or more preliminary interpretations of the utterance. In one implementation, to generate the preliminary interpretations of the utterance, the ASR 120 may be configured to recognize one or more syllables, words, phrases, or other acoustic characteristics from the utterance using one or more dynamic recognition grammars and/or acoustic models. For example, in one implementation, the ASR 120 may use the dynamic recognition grammars and/or the acoustic models

10

to recognize a stream of phonemes from the utterance based on phonetic dictation techniques, as described in U.S. patent application Ser. No. 11/513,269, entitled "Dynamic Speech Sharpening," which issued as U.S. Pat. No. 7,634,409 on Dec. 15, 2009, the contents of which are hereby incorporated by reference in their entirety. In addition, the dynamic recognition grammars and/or the acoustic models may include unstressed central vowels (e.g., "schwa"), which may reduce a search space for recognizing the stream of phonemes for the utterance.

Furthermore, in one implementation, the ASR 120 may be configured as a multi-pass speech recognition engine 120, as described in U.S. patent application Ser. No. 11/197,504, entitled "Systems and Methods for Responding to Natural Language Speech Utterance," which issued as U.S. Pat. No. 7,640,160 on Dec. 29, 2009, the contents of which are hereby incorporated by reference in their entirety. The multi-pass speech recognition 120 may be configured to initially invoke a primary speech recognition engine to generate a first transcription of the utterance, and further to optionally subsequently invoke one or more secondary speech recognition engines to generate one or more secondary transcriptions of the utterance. In one implementation, the first transcription may be generated using a large list dictation grammar, while the secondary transcriptions may be generated using virtual dictation grammars having decoy words for out-of-vocabulary words, reduced vocabularies derived from a conversation history, or other dynamic recognition grammars. For example, in one implementation, if a confidence level for the first transcription does not meet or exceed a threshold, the secondary speech recognition engines may be invoked to sharpen the interpretation of the primary speech recognition engine. It will be apparent, however, that the multi-pass speech recognition engine 120 may interpret the utterance using any suitable combination of techniques that results in a preliminary interpretation derived from a plurality of transcription passes for the utterance (e.g., the secondary speech recognition engines may be invoked regardless of the confidence level for the first transcription, or the primary speech recognition engine and/or the secondary speech recognition engines may employ recognition grammars that are identical or optimized for a particular interpretation context, etc.).

Accordingly, in one implementation, the dynamic recognition grammars used in the ASR 120 may be optimized for different languages, contexts, domains, memory constraints, and/or other suitable criteria. For example, in one implementation, the voice-enabled device 100 may include one or more applications 134 that provide content or services for a particular context or domain, such as a navigation application 134. As such, in response to the ASR 120 determining navigation as the most likely context for the utterance, the dynamic recognition grammars may be optimized for various physical, temporal, directional, or other geographical characteristics (e.g., as described in co-pending U.S. patent application Ser. No. 11/954,064, entitled "System and Method for Providing a Natural Language Voice User Interface in an Integrated Voice Navigation Services Environment," filed Dec. 11, 2007, the contents of which are hereby incorporated by reference in their entirety). In another example, an utterance containing the word "traffic" may be subject to different interpretations depending on whether the user intended a navigation context (i.e., traffic on roads), a music context (i.e., the 1960's rock band), or a movie context (i.e., the Steven Soderbergh film). Accordingly, the recognition grammars used in the ASR 120 may be dynamically adapted to optimize accurate recognition for any given

US 9,502,025 B2

11

utterance (e.g., in response to incorrectly interpreting an utterance to contain a particular word or phrase, the incorrect interpretation may be removed from the recognition grammar to prevent repeating the incorrect interpretation).

In one implementation, in response to the ASR **120** generating the preliminary interpretations of the utterance included in the multi-modal interaction using one or more of the techniques described above, the ASR **120** may provide the preliminary interpretations to the conversational language processor **130**. The conversational language processor **130** may generally include various natural language processing components, which may be configured to model human-to-human conversations or interactions. Thus, the conversational language processor **130** may invoke one or more of the natural language processing components to further analyze the preliminary interpretations of the utterance and any accompanying non-voice interactions to determine the intent of the multi-modal interactions received at the voice-enabled device **100**.

In one implementation, the conversational language processor **120** may invoke an intent determination engine **130a** configured to determine the intent of the multi-modal interactions received at the voice-enabled device **100**. In one implementation, the intent determination engine **130a** may invoke a knowledge-enhanced speech recognition engine that provides long-term and short-term semantic knowledge for determining the intent, as described in co-pending U.S. patent application Ser. No. 11/212,693, entitled "Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions," filed Aug. 29, 2005, the contents of which are hereby incorporated by reference in their entirety. For example, in one implementation, the semantic knowledge may be based on a personalized cognitive model derived from one or more prior interactions with the user, a general cognitive model derived from one or more prior interactions with various different users, and/or an environmental cognitive model derived from an environment associated with the user, the voice-enabled device **100**, and/or the voice services environment (e.g., ambient noise characteristics, location sensitive information, etc.).

Furthermore, the intent determination engine **132a** may invoke a context tracking engine **132d** to determine the context for the multi-modal interactions. For example, any context derived from the natural language utterance and/or the non-voice interactions in the multi-modal interactions may be pushed to a context stack associated with the context tracking engine **132d**, wherein the context stack may include various entries that may be weighted or otherwise ranked according to one or more contexts identified from the cognitive models and the context for the current multi-modal interactions. As such, the context tracking engine **132d** may determine one or more entries in the context stack that match information associated with the current multi-modal interactions to determine a most likely context for the current multi-modal interactions. The context tracking engine **132d** may then provide the most likely context to the intent determination engine **132a**, which may determine the intent of the multi-modal interactions in view of the most likely context.

In addition, based on the most likely context, the intent determination engine **132a** may reference the constellation model **132b** to determine whether to invoke any of the various systems or multi-modal devices in the voice services environment. For example, as noted above, the constellation model **132b** may provide intent determination capabilities, domain knowledge, semantic knowledge, cognitive models, and other information available through the various systems

12

and multi-modal devices. As such, the intent determination engine **132a** may reference the constellation model **132b** to determine whether one or more of the other systems and/or multi-modal devices should be engaged to participate in determining the intent of the multi-modal interactions. For example, in response to the constellation model **132b** indicating that one or more of the other systems and/or multi-modal devices have natural language processing capabilities optimized for the most likely context, the intent determination engine **132a** may forward information relating to the multi-modal interactions to such systems and/or multi-modal devices, which may then determine the intent of the multi-modal interactions and return the intent determination to the voice-enabled device **100**.

In one implementation, the conversational language processor **130** may be configured to engage the user in one or more cooperative conversations to resolve the intent or otherwise process the multi-modal interactions, as described in co-pending U.S. patent application Ser. No. 11/580,926, entitled "System and Method for a Cooperative Conversational Voice User Interface," filed Oct. 16, 2006, the contents of which are hereby incorporated by reference in their entirety. In particular, the conversational language processor **130** may generally identify a conversational goal for the multi-modal interactions, wherein the conversational goal may be identifying from analyzing the utterance, the non-voice interactions, the most likely context, and/or the determined intent. As such, the conversational goal identified for the multi-modal interactions may generally control the cooperative conversation between the conversational language processor **130** and the user. For example, the conversational language processor **130** may generally engage the user in one or more query conversations, didactic conversations, and/or exploratory conversations to resolve or otherwise process the multi-modal interactions.

In particular, the conversational language processor **130** may engage the user in a query conversation in response to identifying that the conversational goal relates to retrieving discrete information or performing a particular function. Thus, in a cooperative query conversation, the user may lead the conversation towards achieving the particular conversational goal, while the conversational language processor **130** may initiate one or more queries, tasks, commands, or other requests to achieve the goal and thereby support the user in the conversation. In response to ambiguity or uncertainty in the intent of the multi-modal interaction, the conversational language processor **130** may engage the user in a didactic conversation to resolve the ambiguity or uncertainty (e.g., where noise or malapropisms interfere with interpreting the utterance, multiple likely contexts cannot be disambiguated, etc.). As such, in a cooperative didactic conversation, the conversational language processor **130** may lead the conversation to clarify the intent of the multi-modal interaction (e.g., generating feedback provided through an output device **114**), while the user may regulate the conversation and provide additional multi-modal interactions to clarify the intent. In response to determining the intent of the multi-modal interactions with suitable confidence, with the intent indicating an ambiguous or uncertain goal, the conversational language processor **130** may engage the user in an exploratory conversation to resolve the goal. In a cooperative exploratory conversation, the conversational language processor **130** and the user may share leader and supporter roles, wherein the ambiguous or uncertain goal may be improvised or refined over a course of the conversation.

US 9,502,025 B2

13

Thus, the conversational language processor **130** may generally engage in one or more cooperative conversations to determine the intent and resolve a particular goal for the multi-modal interactions received at the voice-enabled device **100**. The conversational language processor **130** may then initiate one or more queries, tasks, commands, or other requests in furtherance of the intent and the goal determined for the multi-modal interactions. For example, in one implementation, the conversational language processor **130** may invoke one or more agents **132c** having capabilities for processing requests in a particular domain or application **134**, a voice search engine **132f** having capabilities for retrieving information requested in the multi-modal interactions (e.g., from one or more data repositories **136**, networks, or other information sources coupled to the voice-enabled device **100**), or one or more other systems or multi-modal devices having suitable processing capabilities for furthering the intent and the goal for the multi-modal interactions (e.g., as determined from the constellation model **132b**).

Additionally, in one implementation, the conversational language processor **130** may invoke an advertising application **134** in relation to the queries, tasks, commands, or other requests initiated to process the multi-modal interactions, wherein the advertising application **134** may be configured, to select one or more advertisements that may be relevant to the intent and/or the goal for the multi-modal interactions, as described in co-pending U.S. patent application Ser. No. 11/671,526, entitled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input," filed Feb. 6, 2007, the contents of which are hereby incorporated by reference in their entirety.

In one implementation, in response to receiving results from any suitable combination of queries, tasks, commands, or other requests processed for the multi-modal interactions, the conversational language processor **130** may format the results for presentation to the user through the output device **114**. For example, the results may be formatted into a natural language utterance that can be converted into an electronic signal and provided to the user through a speaker coupled to the output device **114**, or the results may be visually presented on a display coupled to the output device **114**, or in any other suitable manner (e.g., the results may indicate whether a particular task or command was successfully performed, or the results may include information retrieved in response to one or more queries, or the results may include a request to frame a subsequent multi-modal interaction if the results are ambiguous or otherwise incomplete, etc.).

Furthermore, in one implementation, the conversational language processor **130** may include a misrecognition engine **132e** configured to determine whether the conversational language processor **130** incorrectly determined the intent for the multi-modal interactions. In one implementation, the misrecognition engine **132e** may determine that the conversational language processor **130** incorrectly determined the intent in response to one or more subsequent multi-modal interactions provided proximately in time to the prior multi-modal interactions, as described in U.S. patent application Ser. No. 11/200,164, entitled "System and Method of Supporting Adaptive Misrecognition in Conversational Speech," which issued as U.S. Pat. No. 7,620,549 on Nov. 17, 2009, the contents of which are hereby incorporated by reference in their entirety. For example, the misrecognition engine **132e** may monitor for one or more subsequent multi-modal interactions that include a stop

14

word, override a current request, or otherwise indicate an unrecognized or misrecognized event. The misrecognition engine **132e** may then determine one or more tuning parameters for various components associated with the ASR **120** and/or the conversational language processor **130** to improve subsequent interpretations.

Accordingly, as described in further detail above, the voice-enabled device **100** may generally include various natural language processing components and capabilities that may be used for hybrid processing in the natural language voice services environment. In particular, the voice-enabled device **100** may be configured to determine the intent for various multi-modal interactions that include any suitable combination of natural language utterances and/or non-voice interactions and process one or more queries, tasks, commands, or other requests based on the determined intent. Furthermore, as noted above and as will be described in greater detail below, one or more other systems and/or multi-modal devices may participate in determining the intent and processing the queries, tasks, commands, or other requests for the multi-modal interactions to provide a hybrid processing methodology, wherein the voice-enabled device **100** and the various other systems and multi-modal devices may each perform partial processing to determine the intent and otherwise process the multi-modal interactions in a cooperative manner. For example, in one implementation, hybrid processing in the natural language voice services environment may include one or more techniques described in U.S. Provisional Patent Application Ser. No. 61/259,827, entitled "System and Method for Hybrid Processing in a Natural Language Voice Services Environment," filed on Nov. 10, 2009, the contents of which are hereby incorporated by reference in their entirety.

According to one aspect of the invention, FIG. 2 illustrates a block diagram of an exemplary system for hybrid processing in a natural language voice service environment. In particular, the system illustrated in FIG. 2 may generally include a voice-enabled client device **210** similar to the voice-enabled device described above in relation to FIG. 1. For example, the voice-enabled client device **210** may include any suitable combination of input and output devices **215a** respectively arranged to receive natural language multi-modal interactions and provide responses to the natural language multi-modal interactions. In addition, the voice-enabled client device **210** may include an Automatic Speech Recognizer (ASR) **220a** configured to generate one or more preliminary interpretations of natural language utterances received at the input device **215a**, and further configured to provide the preliminary interpretations to a conversational language processor **230a**.

In one implementation, the conversational language processor **230a** on the voice-enabled client device **210** may include one or more natural language processing components, which may be invoked to determine an intent for the multi-modal interactions received at the voice-enabled client device **210**. The conversational language processor **230a** may then initiate one or more queries, tasks, commands, or other requests to resolve the determined intent. For example, the conversational language processor **230a** may invoke one or more applications **234a** to process requests in a particular domain, query one or more data repositories **236a** to retrieve information requested in the multi-modal interactions, or otherwise engage in one or more cooperative conversations with a user of the voice-enabled client device **210** to resolve the determined intent. Furthermore, as noted above in connection with FIG. 1, the voice-enabled client device **210** may also cooperate with one or more other systems or

US 9,502,025 B2

15

multi-modal devices having suitable processing capabilities for initiating queries, tasks, commands, or other requests to resolve the intent of the multi-modal interactions.

In particular, to cooperate with the other systems or multi-modal devices in the hybrid processing environment, the voice-enabled client device **210** may use a messaging interface **250a** to communicate with a virtual router **260**, wherein the messaging interface **250a** may generally include a light client (or thin client) that provides a mechanism for the voice-enabled client device **210** to transmit input to and receive output from the virtual router **260**. In addition, the virtual router **260** may further include a messaging interface **250b** providing a mechanism for communicating with one or more additional voice-enabled devices **270a-n**, one or more non-voice devices **280a-n**, and a voice-enabled server **240**. Furthermore, although FIG. 2 illustrates messaging interface **250a** and messaging interface **250b** as components that are distinct from the devices to which they are communicatively coupled, it will be apparent that such illustration is for ease of description only, as the messaging interfaces **250a-b** may be provided as on-board components that execute on the various devices illustrated in FIG. 2 to facilitate communication among the various devices in the hybrid processing environment.

For example, in one implementation, the messaging interface **250a** that executes on the voice-enabled client device **210** may transmit input from the voice-enabled client device **210** to the virtual router **260** within one or more XML messages, wherein the input may include encoded audio corresponding to natural language utterances, preliminary interpretations of the natural language utterances, data corresponding to multi-touch gestures, point of focus or attention focus selections, and/or other multi-modal interactions. In one implementation, the virtual router **260** may then further process the input using a conversational language processor **230c** having capabilities for speech recognition, intent determination, adaptive misrecognition, and/or other natural language processing. Furthermore, the conversational language processor **230c** may include knowledge relating to content, services, applications, natural language processing capabilities, and other features available through the various devices in the hybrid processing environment.

As such, in one implementation, the virtual router **260** may further communicate with the voice-enabled devices **270**, the non-voice devices **280**, and/or the voice-enabled server **240** through the messaging interface **250b** to coordinate processing for the input received from the voice-enabled client device **210**. For example, based on the knowledge relating to the features and capabilities of the various devices in the hybrid processing environment, the virtual router **260** may identify one or more of the devices that have suitable features and/or capabilities for resolving the intent of the input received from the voice-enabled client device **210**. The virtual router **260** may then forward one or more components of the input to the identified devices through respective messaging interfaces **250b**, wherein the identified devices may be invoked to perform any suitable processing for the components of the input forwarded from the virtual router **260**. In one implementation, the identified devices may then return any results of the processing to the virtual router **260** through the respective messaging interfaces **250b**, wherein the virtual router **260** may collate the results of the processing and return the results to the voice-enabled client device **210** through the messaging interface **250a**.

Accordingly, the virtual router **260** may communicate with any of the devices available in the hybrid processing

16

environment through messaging interfaces **250a-b** to coordinate cooperative hybrid processing for multi-modal interactions or other natural language inputs received from the voice-enabled client device **210**. For example, in one implementation, the cooperative hybrid processing may be used to enhance performance in embedded processing architectures in which the voice-enabled client device **210** includes a constrained amount of resources (e.g., the voice-enabled client device **210** may be a mobile device having a limited amount of internal memory or other dedicated resources for natural language processing). As such, when the voice-enabled client device **210** has an embedded processing architecture, one or more components of the voice-enabled client device **210** may be configured to optimize efficiency of on-board natural language processing to reduce or eliminate bottlenecks, lengthy response times, or degradations in performance.

For example, in one implementation, optimizing the efficiency of the on-board natural language processing may include configuring the ASR **220a** to use a virtual dictation grammar having decoy words for out-of-vocabulary words, reduced vocabularies derived from a conversation history, or other dynamic recognition grammars (e.g., grammars optimized for particular languages, contexts, domains, memory constraints, and/or other suitable criteria). In another example, the on-board applications **234a** and/or data repositories **236a** may be associated with an embedded application suite providing particular features and capabilities for the voice-enabled client device **210**. For example, the voice-enabled client device **210** may be embedded within an automotive telematics system, a personal navigation device, a global positioning system, a mobile phone, or another device in which users often request location-based services. Thus, in such circumstances, the on-board applications **234a** and the data repositories **236a** in the embedded application suite may be optimized to provide certain location-based services that can be efficiently processed on-board (e.g., destination entry, navigation, map control, music search, hands-free dialing, etc.).

Furthermore, although the components of the voice-enabled client device **210** may be optimized for efficiency in embedded architectures, a user may nonetheless request any suitable content, services, applications, and/or other features available in the hybrid processing environment, and the other devices in the hybrid processing environment may collectively provide natural language processing capabilities to supplement the embedded natural language processing capabilities for the voice-enabled client device **210**. For example, the voice-enabled client device **210** may perform preliminary processing for a particular multi-modal interaction using the embedded natural language processing capabilities (e.g., the on-board ASR **220a** may perform advanced virtual dictation to partially transcribe an utterance in the multi-modal interaction, the on-board conversational language processor **230a** may determine a preliminary intent of the multi-modal interaction, etc.), wherein results of the preliminary processing may be provided to the virtual router **260** for further processing.

In one implementation, the voice-enabled client device **210** may also communicate input corresponding to the multi-modal interaction to the virtual router **260** in response to determining that on-board capabilities cannot suitably interpret the interaction (e.g., if a confidence level for a partial transcription does not satisfy a particular threshold), or in response to determining that the interaction should be processed off-board (e.g., if a preliminary interpretation indicates that the interaction relates to a local search request

US 9,502,025 B2

17

requiring large computations to be performed on the voice-enabled server **240**). As such, the virtual router **260** may capture the input received from the voice-enabled client device **210** and coordinate further processing among the voice-enabled devices **270** and the voice-enabled server **240** that provide natural language processing capabilities in addition to the non-voice devices **280** that provide capabilities for retrieving data or executing tasks. Furthermore, in response to the virtual router **260** invoking one or more of the voice-enabled devices **270**, the input provided to the voice-enabled devices **270** may be optimized to suit the processing requested from the invoked voice-enabled devices **270** (e.g., to avoid over-taxing processing resources, a particular voice-enabled device **270** may be provided a partial transcription or a preliminary interpretation and resolve the intent for a given context or domain).

Alternatively, in response to the virtual router **260** invoking the voice-enabled server **240**, the input provided to the voice-enabled devices **270** may further include encoded audio corresponding to natural language utterances and any other data associated with the multi-modal interaction. In particular, as shown in FIG. 2, the voice-enabled server **240** may have a natural language processing architecture similar to the voice-enabled client device **210**, except that the voice-enabled server **240** may include substantial processing resources that obviate constraints that the voice-enabled client device **210** may be subject to. Thus, when the voice-enabled server **240** cooperates in the hybrid processing for the multi-modal interaction, the encoded audio corresponding to the natural language utterances and the other data associated with the multi-modal interaction may be provided to the voice-enabled server **240** to maximize a likelihood of the voice-enabled server **240** correctly determining the intent of the multi-modal interaction (e.g., the ASR **220b** may perform multi-pass speech recognition to generate an accurate transcription for the natural language utterance, the conversational language processor **230b** may arbitrate among intent determinations performed in any number of different contexts or domains, etc.). Accordingly, in summary, the hybrid processing techniques performed in the environment illustrated in FIG. 2 may generally include various different devices, which may or may not include natural language capabilities, cooperatively determining the intent of a particular multi-modal interaction and taking action to resolve the intent.

Although the cooperative hybrid processing techniques described above have been particularly described in the context of an embedded processing architecture, such techniques are not necessarily limited to embedded processing architectures. In particular, the same techniques may be applied in any suitable voice services environment having various devices that can cooperate to initiate queries, tasks, commands, or other requests to resolve the intent of multi-modal interactions. Furthermore, in one implementation, the voice-enabled client device **210** may include a suitable amount of memory or other resources that can be dedicated to natural language processing (e.g., the voice-enabled client device **210** may be a desktop computer or other device that can process natural language without substantially degraded performance). In such circumstances, one or more of the components of the voice-enabled client device **210** may be configured to optimize the on-board natural language processing in a manner that could otherwise cause bottlenecks, lengthy response times, or degradations in performance in an embedded architecture. For example, in one implementation, optimizing the on-board natural language processing may include configuring the ASR **220a** to use a large list

18

dictation grammar in addition to and/or instead of the virtual dictation grammar used in embedded processing architectures.

Nonetheless, as will be described in greater detail below in connection with FIGS. 3-5, the cooperative hybrid processing techniques may be substantially similar regardless of whether the voice-enabled client device **210** has an embedded or non-embedded architecture. In particular, regardless of the architecture for the voice-enabled client device **210**, cooperative hybrid processing may include the voice-enabled client device **210** optionally performing preliminary processing for a natural language multi-modal interaction and communicating input corresponding to the multi-modal interaction to the virtual router **260** for further processing through the messaging interface **250a**. Alternatively (or additionally), the cooperative hybrid processing may include the virtual router **260** coordinating the further processing for the input among the various devices in the hybrid environment through messaging interface **250b**, and subsequently returning any results of the processing to the voice-enabled client device **210** through messaging interface **250a**.

According to various aspects of the invention, FIG. 3 illustrates a flow diagram of an exemplary method for initializing various devices that cooperate to perform hybrid processing in a natural language voice services environment. In particular, as noted above, the hybrid processing environment may generally include communication among various different devices that may cooperatively process natural language multi-modal interactions. For example, in one implementation, the various devices in the hybrid processing environment may include a virtual router having one or more messaging interfaces for communicating with one or more voice-enabled devices, one or more non-voice devices, and/or a voice-enabled server. As such, in one implementation, the method illustrated in FIG. 3 may be used to initialize communication in the hybrid processing environment to enable subsequent cooperative processing for one or more natural language multi-modal interactions received at any particular device in the hybrid processing environment.

In one implementation, the various devices in the hybrid processing environment may be configured to continually listen or otherwise monitor respective input devices to determine whether a natural language multi-modal interaction has occurred. As such, the method illustrated in FIG. 3 may be used to calibrate, synchronize, or otherwise initialize the various devices that continually listen for the natural language multi-modal interactions. For example, as described above in connection with FIG. 2, the virtual router, the voice-enabled devices, the non-voice devices, the voice-enabled server, and/or other devices in the hybrid processing environment may be configured to provide various different capabilities or services, wherein the initialization method illustrated in FIG. 3 may be used to ensure that the hybrid processing environment obtains a suitable signal to process any particular natural language multi-modal interaction and appropriately invoke one or more of the devices to cooperatively process the natural language multi-modal interaction. Furthermore, the method illustrated in FIG. 3 and described herein may be invoked to register the various devices in the hybrid processing environment, register new devices added to the hybrid processing environment, publish domains, services, intent determination capabilities, and/or other features supported on the registered devices, synchronize local timing for the registered devices, and/or initialize any other suitable aspect of the devices in the hybrid processing environment.

US 9,502,025 B2

19

In one implementation, initializing the various devices in the hybrid processing environment may include an operation 310, wherein a device listener may be established for each of the devices in the hybrid processing environment. The device listeners established in operation 310 may generally include any suitable combination of instructions, firmware, or other routines that can be executed on the various devices to determine capabilities, features, supported domains, or other information associated with the devices. For example, in one implementation, the device listeners established in operation 310 may be configured to communicate with the respective devices using the Universal Plug and Play protocol designed for ancillary computer devices, although it will be apparent that any appropriate mechanism for communicating with the various devices may be suitably substituted.

In response to establishing the device listeners for each device registered in the hybrid processing environment (or in response to establishing device listeners for any device newly registered in the hybrid processing environment), the device listeners may then be synchronized in an operation 320. In particular, each of the registered devices may have an internal clock or other timing mechanism that indicates local timing for an incoming natural language multi-modal interaction, wherein operation 320 may be used to synchronize the device listeners established in operation 310 according to the internal clocks or timing mechanisms for the respective devices. Thus, in one implementation, synchronizing the device listeners in operation 320 may include each device listener publishing information relating to the internal clock or local timing for the respective device. For example, the device listeners may publish the information relating to the internal clock or local timing to the virtual router, whereby the virtual router may subsequently coordinate cooperative hybrid processing for natural language multi-modal interactions received at one or more of the devices in the hybrid processing environment. It will be apparent, however, that the information relating to the internal clock or local timing for the various devices in the hybrid processing environment may be further published to the other voice-enabled devices, the other non-voice devices, the voice-enabled server, and/or any other suitable device that may participate in cooperative processing for natural language multi-modal interactions provided to the hybrid processing environment.

In one implementation, in response to establishing and synchronizing the device listeners for the various devices registered in the hybrid processing environment, the device listeners may continually listen or otherwise monitor respective devices on the respective registered devices in an operation 330 to detect information relating to one or more natural language multi-modal interactions. For example, the device listeners may be configured to detect occurrences of the natural language multi-modal interactions in response to detecting an incoming natural language utterance, a point of focus or attention focus selection associated with an incoming natural language utterance, and/or another interaction or sequence of interactions that relates to an incoming natural language multi-modal interaction. In addition, operation 330 may further include the appropriate device listeners capturing the natural language utterance and/or related non-voice device interactions that relate to the natural language utterance.

In one implementation, the captured natural language utterance and related non-voice device interactions may then be analyzed in an operation 340 to manage subsequent cooperative processing in the hybrid processing environ-

20

ment. In one implementation, for example, operation 340 may determine whether one device listener or multiple device listeners captured information relating to the natural language multi-modal interaction detected in operation 330. In particular, as noted above, the hybrid processing environment may generally include various different devices that cooperate to process natural language multi-modal interactions, whereby the information relating to the natural language multi-modal interaction may be provided to one or a plurality of the devices in the hybrid processing environment. As such, operation 340 may determine whether one device listener or multiple device listeners captured the information relating to the natural language multi-modal interaction in order to determine whether the hybrid processing environment needs to synchronize signals among various device listeners that captured information relating to the multi-modal interaction.

For example, a user interacting with the hybrid processing environment may view a web page presented on a non-voice display device and provide a natural language multi-modal interaction that requests more information about purchasing a product displayed on the web page. The user may then select text on the web page containing the product name using a mouse, keyboard, or other non-voice input device and provide a natural language utterance to a microphone or other voice-enabled device such as "Is this available on Amazon.com?" In this example, a device listener associated with the non-voice display device may detect the text selection for the product name in operation 330, and a device listener associated with the voice-enabled device may further detect the natural language utterance inquiring about the availability of the product in operation 330. Furthermore, in one implementation, the user may be within a suitable range of multiple voice-enabled devices, which may result in multiple device listeners capturing different signals corresponding to the natural language utterance (e.g., the interaction may occur within range of a voice-enabled mobile phone, a voice-enabled telematics device, and/or other voice-enabled devices, depending on the arrangement and configuration of the various devices in the hybrid processing environment).

Accordingly, as will be described in greater detail herein, a sequence of operations that synchronizes different signals relating to the multi-modal interaction received at the multiple device listeners may be initiated in response to operation 340 determining that multiple device listeners captured information relating to the natural language multi-modal interaction. On the other hand, in response to operation 340 determining that only one device listener captured information relating to the natural language multi-modal interaction, the natural language multi-modal interaction may be processed in an operation 390 without executing the sequence of operations that synchronizes different signals (i.e., the one device listener provides all of the input information relating to the multi-modal interaction, such that hybrid processing for the interaction may be initiated in operation 390 without synchronizing different input signals). However, in one implementation, the sequence of synchronization operations may also be initiated in response to one device listener capturing a natural language utterance and one or more non-voice interactions to align different signals relating to the natural language multi-modal interaction, as described in greater detail herein.

As described above, each device listener that receives an input relating to the natural language multi-modal interaction detected in operation 330 may have an internal clock or other local timing mechanism. As such, in response to

US 9,502,025 B2

21

determining that one or more device listeners captured different signals relating to the natural language multi-modal interaction in operation 340, the sequence of synchronization operations for the different signals may be initiated in an operation 350. In particular, operation 350 may include the one or more device listeners determining local timing information for the respective signals based on the internal clock or other local timing mechanism associated with the respective device listeners, wherein the local timing information determined for the respective signals may then be synchronized.

For example, in one implementation, synchronizing the local timing information for the respective signals may be initiated in an operation 360. In particular, operation 360 may generally include notifying each device listener that received an input relating to the multi-modal interaction of the local timing information determined for each respective signal. For example, in one implementation, each device listener may provide local timing information for a respective signal to the virtual router, and the virtual router may then provide the local timing information for all of the signals to each device listener. As such, in one implementation, operation 360 may result in each device listener receiving a notification that includes local timing information for each of the different signals that relate to the natural language multi-modal interaction detected in operation 330. Alternatively (or additionally), the virtual router may collect the local timing information for each of the different signals from each of the device listeners and further synchronize the local timing information for the different signals to enable hybrid processing for the natural language multi-modal interaction.

In one implementation, any particular natural language multi-modal interaction may include at least a natural language utterance, and may further include one or more additional device interactions relating to the natural language utterance. As noted above, the utterance may generally be received prior to, contemporaneously with, or subsequent to the additional device interactions. As such, the local timing information for the different signals may be synchronized in an operation 370 to enable hybrid processing for the natural language multi-modal interaction. In particular, operation 370 may include aligning the local timing information for one or more signals corresponding to the natural language utterance and/or one or more signals corresponding to any additional device interactions that relate to the natural language utterance. In addition, operation 370 may further include aligning the local timing information for the natural language utterance signals with the signals corresponding to the additional device interactions.

Thus, in matching the utterance signals and the non-voice device interaction signals, any devices that participate in the hybrid processing for the natural language multi-modal interaction may be provided with voice components and/or non-voice components that have been aligned with one another. For example, in one implementation, operation 370 may be executed on the virtual router, which may then provide the aligned timing information to any other device that may be invoked in the hybrid processing. Alternatively (or additionally), one or more of the other devices that participate in the hybrid processing may locally align the timing information (e.g., in response to the virtual router invoking the voice-enabled server in the hybrid processing, resources associated with the voice-enabled server may be employed to align the timing information and preserve communication bandwidth at the virtual router).

22

Furthermore, in one implementation, the virtual router and/or other devices in the hybrid processing environment may analyze the signals corresponding to the natural language utterance in an operation 380 to select the cleanest sample for further processing. In particular, as noted above, the virtual router may include a messaging interface for receiving an encoded audio sample corresponding to the natural language utterance from one or more of the voice-enabled devices. For example, the audio sample received at the virtual router may include the natural language utterance encoded in the MPEG-1 Audio Layer 3 (MP3) format or another lossy format to preserve communication bandwidth in the hybrid processing environment. However, it will be apparent that the audio sample may alternatively (or additionally) be encoded using the Free Lossless Audio Codec (FLAC) format or another lossless format in response to the hybrid processing environment having sufficient communication bandwidth for processing lossless audio that may provide a better sample of the natural language utterance.

Regardless of whether the audio sample has been encoded in a lossy or lossless format, the signal corresponding to the natural language utterance that provides the cleanest sample may be selected in operation 380. For example, one voice-enabled device may be in a noisy environment or otherwise associated with conditions that interfere with generating a clean audio sample, while another voice-enabled device may include a microphone array or be configured to employ techniques that maximize fidelity of encoded speech. As such, in response to multiple signals corresponding to the natural language utterance being received in operation 330, the cleanest signal may be selected in operation 380 and hybrid processing for the natural language utterance may then be initiated in an operation 390.

Accordingly, the synchronization and initialization techniques illustrated in FIG. 3 and described herein may ensure that the hybrid processing environment synchronizes each of the signals corresponding to the natural language multi-modal interaction and generates an input for further processing in operation 390 most likely to result in a correct intent determination. Furthermore, in synchronizing the signals and selecting the cleanest audio sample for the further processing in operation 390, the techniques illustrated in FIG. 3 and described herein may ensure that none of the devices in the hybrid processing environment take action on a natural language multi-modal interaction until the appropriate signals to be used in operation 390 have been identified. As such, hybrid processing for the natural language multi-modal interaction may be initiated in operation 390, as described in further detail herein.

According to one aspect of the invention, FIG. 4 illustrates a flow diagram of an exemplary method for performing hybrid processing at one or more client devices in a natural language voice services environment. In particular, as will be described in greater detail below with reference to FIG. 5, the one or more client devices may perform the hybrid processing in cooperation with a virtual router through a messaging interface that communicatively couples the client devices and the virtual router. For example, in one implementation, the messaging interface may generally include a light client (or thin client) that provides a mechanism for the client devices to transmit input relating to a natural language multi-modal interaction to the virtual router, and that further provides a mechanism for the client devices to receive output relating to the natural language multi-modal interaction from the virtual router.

For example, in one implementation, the hybrid processing at the client devices may be initiated in response to one

US 9,502,025 B2

23

or more of the client devices receiving a natural language multi-modal interaction in an operation **410**. In particular, the natural language multi-modal interaction may generally include a natural language utterance received at a microphone or other voice-enabled input device coupled to the client device that received the natural language multi-modal interaction, and may further include one or more other additional input modalities that relate to the natural language utterance (e.g., text selections, button presses, multi-touch gestures, etc.). As such, the natural language multi-modal interaction received in operation **410** may include one or more queries, commands, or other requests provided to the client device, wherein the hybrid processing for the natural language multi-modal interaction may then be initiated in an operation **420**.

As described in greater detail above, the natural language voice services environment may generally include one or more voice-enabled client devices, one or more non-voice devices, a voice-enabled server, and a virtual router arranged to communicate with each of the voice-enabled client devices, the non-voice devices, and the voice-enabled server. In one implementation, the virtual router may therefore coordinate the hybrid processing for the natural language multi-modal interaction among the voice-enabled client devices, the non-voice devices, and the voice-enabled server. As such, the hybrid processing techniques described herein may generally refer to the virtual router coordinating cooperative processing for the natural language multi-modal interaction in a manner that involves resolving an intent of the natural language multi-modal interaction in multiple stages.

In particular, as described above in connection with FIG. 3, the various devices that cooperate to perform the hybrid processing may be initialized to enable the cooperative processing for the natural language multi-modal interaction. As such, in one implementation, in response to initializing the various devices, each of the client devices that received an input relating to the natural language multi-modal interaction may perform initial processing for the respective input in an operation **420**. For example, in one implementation, a client device that received the natural language utterance included in the multi-modal interaction may perform initial processing in operation **420** that includes encoding an audio sample corresponding to the utterance, partially or completely transcribing the utterance, determining a preliminary intent for the utterance, or performing any other suitable preliminary processing for the utterance. In addition, the initial processing in operation **420** may also be performed at a client device that received one or more of the additional input modalities relating to the utterance. For example, the initial processing performed in operation **420** for the additional input modalities may include identifying selected text, selected points of focus or attention focus, or generating any other suitable data that can be used to further interpret the utterance. In one implementation, an operation **430** may then include determining whether the hybrid processing environment has been configured to automatically route inputs relating to the natural language multi-modal interaction to the virtual router.

For example, in one implementation, operation **430** may determine that automatic routing has been configured to occur in response to multiple client devices receiving the natural language utterance included in the multi-modal interaction in operation **410**. In this example, the initial processing performed in operation **420** may include the multiple client devices encoding respective audio samples corresponding to the utterance, wherein messages that

24

include the encoded audio samples may then be sent to the virtual router in an operation **460**. The virtual router may then select one of the encoded audio samples that provides a cleanest signal and coordinate subsequent hybrid processing for the natural language multi-modal interaction, as will be described in greater detail below with reference to FIG. 5. In another example, operation **430** may determine that automatic routing has been configured to occur in response to the initial processing resulting in a determination that the multi-modal interaction relates to a request that may be best suited for processing on the voice-enabled server (e.g., the request may relate to a location-based search query or another command or task that requires resources managed on the voice-enabled server, content, applications, domains, or other information that resides on one or more devices other than the client device that received the request, etc.). However, it will be apparent that the hybrid processing environment may be configured for automatic routing in response to other conditions and/or regardless of whether any attendant conditions exist, as appropriate.

In one implementation, in response to the virtual router coordinating the hybrid processing for the natural language multi-modal interaction, the virtual router may provide results of the hybrid processing to the client device in an operation **470**. For example, the results provided to the client device in operation **470** may include a final intent determination for the natural language multi-modal interaction, information requested in the interaction, data generated in response to executing a command or task requested in the interaction, and/or other results that enable the client device to complete processing for the natural language request in operation **480**. For example, in one implementation, operation **480** may include the client device executing a query, command, task, or other request based on the final intent determination returned from the virtual router, presenting the requested information returned from the virtual router, confirming that the requested command or task has been executed, and/or performing any additional processing to resolve the natural language request.

Referring back to operation **430**, in response to determining that the conditions that trigger automatic routing have not been satisfied or that automatic routing has otherwise not been configured, the client device may further process the natural language multi-modal interaction in an operation **440**. In one implementation, the further processing in operation **440** may include the client device attempting to determine an intent for the natural language multi-modal interaction using local natural language processing capabilities. For example, the client device may merge any non-voice input modalities included in the multi-modal interaction a transcription for the utterance included in the multi-modal interaction. The conversational language processor on the client device may then determine the intent for the multi-modal interaction utilizing local information relating to context, domains, shared knowledge, criteria values, or other information. The client device may then generate one or more interpretations for the utterance to determine the intent for the multi-modal interaction (e.g., identifying a conversation type, one or more requests contained in the interactions, etc.).

In one implementation, operation **440** may further include determining a confidence level for the intent determination generated on the client device (e.g., the confidence level may be derived in response to whether the client devices includes a multi-pass speech recognition engine, whether the utterance contained any ambiguous words or phrases, whether the intent differs from one context to another, etc.). In one

US 9,502,025 B2

25

implementation, an operation 450 may then determine whether or not to invoke off-board processing depending on the confidence level determined in operation 440. For example, operation 450 may generally include determining whether the intent determined in operation 440 satisfies a particular threshold value that indicates an acceptable confidence level for taking action on the determined intent. As such, in response to the confidence level for the intent determination satisfying the threshold value, operation 450 may determine to not invoke off-board processing. In particular, the confidence level satisfying the threshold value may indicate that the client device has sufficient information to take action on the determined intent, whereby the client device may then process one or more queries, commands, tasks, or other requests to resolve the multi-modal interaction in operation 480.

Alternatively, in response to the confidence level for the intent determination failing to satisfy the threshold value, operation 450 may invoke off-board processing, which may include sending one or more messages to the virtual router in operation 460. The one or more messages may cause the virtual router to invoke additional hybrid processing for the multi-modal interaction in a similar manner as noted above, and as will be described in greater detail herein with reference to FIG. 5.

According to one aspect of the invention, FIG. 5 illustrates a flow diagram of an exemplary method for performing hybrid processing at a virtual router in a natural language voice services environment. In particular, as the virtual router may coordinate the hybrid processing for natural language multi-modal interactions received at one or more client devices. In one implementation, in an operation 510, the virtual router may receive one or more messages relating to a natural language multi-modal interaction received at one or more of the client devices in the voice services environment. For example, the virtual router may include a messaging interface that communicatively couples the virtual router to the client devices and a voice-enabled server, wherein the messaging interface may generally include a light client (or thin client) that provides a mechanism for the virtual router to receive input from one or more the client devices and/or the voice-enabled server, and further to transmit output to one or more the client devices and/or the voice-enabled server. The messages received in operation 510 may generally include any suitable processing results for the multi-modal interactions, whereby the virtual router may coordinate hybrid processing in a manner that includes multiple processing stages that may occur at the virtual router, one or more of the client devices, the voice-enabled server, or any suitable combination thereof.

In one implementation, the virtual router may analyze the messages received in operation 510 to determine whether to invoke the hybrid processing in a peer-to-peer mode. For example, one or more of the messages may include a preliminary intent determination that the virtual router can use to determine whether to invoke one or more of the client devices, the voice-enabled server, or various combinations thereof in order to execute one or more of the multiple processing stages for the multi-modal interaction. In another example, one or more of the messages may include an encoded audio sample that the virtual router forwards to one or more of the various devices in the hybrid processing environment. As such, in one implementation, the virtual router may analyze the messages received in operation 510 to determine whether or not to invoke the voice-enabled server to process the multi-modal interaction (e.g., the messages may include a preliminary intent determination

26

that indicates that the multi-modal interaction includes a location-based request that requires resources residing on the server).

In response to the virtual router determining to invoke the voice-enabled server, the virtual router may forward the messages to the server in an operation 530. In particular, the messages forwarded to the server may generally include the encoded audio corresponding to the natural language utterance and any additional information relating to other input modalities relevant to the utterance. For example, as described in greater detail above with reference to FIG. 2, the voice-enabled server may include various natural language processing components that can suitably determine the intent of the multi-modal interaction, whereby the messages sent to the voice-enabled server may include the encoded audio in order to permit the voice-enabled server to determine the intent independently of any preliminary processing on the client devices that may be inaccurate or incomplete. In response to the voice-enabled server processing the messages received from the virtual router, results of the processing may then be returned to the virtual router in an operation 570. For example, the results may include the intent determination for the natural language multi-modal interaction, results of any queries, commands, tasks, or other requests performed in response to the determined intent, or any other suitable results, as will be apparent.

Alternatively, in response to the virtual router determining to invoke the peer-to-peer mode in operation 520, the virtual router may coordinate the hybrid processing among one or more the client devices, the voice-enabled server, or any suitable combination thereof. For example, in one implementation, the virtual router may determine a context for the natural language multi-modal interaction in an operation 540 and select one or more peer devices based on the determined context in an operation 550. For example, one or more of the client devices may be configured to provide content and/or services in the determined context, whereby the virtual router may forward one or more messages to such devices in an operation 560 in order to request such content and/or services. In another example, the multi-modal interaction may include a compound request that relates to multiple contexts supported on different devices, whereby the virtual router may forward messages to each such device in operation 560 in order to request appropriate content and/or services in the different contexts.

In still another example, the interaction may include a request to be processed on the voice-enabled server, yet the request may require content and/or services that reside on one or more of the client devices (e.g., a location-based query relating to an entry in an address book on one or more of the client devices). As such, the virtual router may generally forward various messages to the selected peer devices in operation 560 to manage the multiple stages in the hybrid processing techniques described herein. For example, the virtual router may send messages to one or more voice-enabled client devices that have intent determination capabilities in a particular context, one or more non-voice client devices that have access to content, services, and/or other resources needed to process the multi-modal interaction, or any appropriate combination thereof. The virtual router may therefore send messages to the client devices and/or the voice-enabled server in operation 560 and receive responsive messages from the client devices and/or the voice-enabled server in operation 570 in any appropriate manner (e.g., in parallel, sequentially, iteratively, etc.). The virtual router may then collate the results received in the responsive messages in operation 580 and return the results

US 9,502,025 B2

27

to one or more of the client devices for any final processing and/or presentation of the results.

According to one aspect of the invention, FIG. 6 illustrates a block diagram of an exemplary system for providing a natural language content dedication service. In particular, as shown in FIG. 6, the system for providing the natural language content dedication service may generally include a voice-enabled client device 610 that can communicate with a content dedication system 665 through a messaging interface 665. In one implementation, the voice-enabled client device 610 may communicate with the content dedication system 665 in a similar manner as described above with reference to FIGS. 2 through 5. For example, the content dedication system 665 may include a virtual router 660 and a voice-enabled server 640 that can service multi-modal natural language requests provided to the voice-enabled client device 610 in a similar manner as described above, and may further include a billing system 638 that may be used to process transactions relating to the content dedication service. Further, although FIG. 6 illustrates the content dedication system 665 as including the virtual router 660, the voice-enabled server 640, and the billing system 638 within one component, such illustration will be understood to be for ease of description only, and that the virtual router 660, the voice-enabled server 640, and the billing system 638 may in fact be arranged within any number of components that can suitably communicate with one another to process multi-modal natural language requests relating to the content dedication service (e.g., the virtual router 660 may communicate with the voice-enabled server 640 through another messaging interface distinct from the messaging interface 650 for communicating with the voice-enabled client device 610, as shown in the exemplary system illustrated in FIG. 2).

In one implementation, the system shown in FIG. 6 may provide the natural language content dedication service to any suitable voice-enabled client device 610 having a suitable combination of input and output devices 615a that can receive natural language multi-modal interactions and provide responses to the natural language multi-modal interactions, wherein the input and output devices 615a may be further arranged to receive any other suitable type of input and provide any other suitable type of output. For example, in one implementation, the voice-enabled client device 610 may comprise a mobile phone that includes a keypad input device 615a, a touch screen input device 615a, or other input mechanisms 615a in addition to any input microphones or other suitable input devices 615a that can receive voice signals. As such, in this example, the mobile phone may further include an output display device 615a in addition to any output microphones or other suitable output devices 615a that can output audible signals. Thus, a user of the voice-enabled client device 610 may be listening to music, watching video, or otherwise interacting with content through the input and output devices 615a and provide a multi-modal natural language request to engage in a transaction to dedicate the music, video, or other content, as will be described in greater detail below. Furthermore, in one implementation, the voice-enabled client device 610 may be included within a hybrid processing environment that may include a plurality of different devices, whereby the content dedication request may relate to content played on a different device from the voice-enabled client device 610, although the content dedication request may relate to any suitable content (i.e., the request need not necessarily relate to played content, as users may provide natural language to request content dedications for any suitable content).

28

Thus, in one implementation, in response to the voice-enabled client device 710 receiving a multi-modal interaction that includes a natural language utterance, the voice-enabled client device 610 may invoke an Automatic Speech Recognizer (ASR) 620a to generate one or more preliminary interpretations of the utterance. The ASR 620a may then provide the preliminary interpretations to a conversational language processor 630a, which may attempt to determine an intent for the multi-modal interaction. In one implementation, to determine the intent for the multi-modal interaction, the conversational language processor 630a may determine a most likely context for the interaction from the preliminary interpretations of the utterance, any accompanying non-speech inputs in the multi-modal interaction that relate to the utterance, contexts associated with prior requests, short-term and long-term shared knowledge, or any other suitable information for interpreting the multi-modal interaction. Thus, in response to the conversational language processor 630a determining that the intent for the multi-modal interaction relates to a content dedication request, a content dedication application 634a may be invoked to resolve the content dedication request.

For example, in one implementation, an initial multi-modal interaction may include the utterance "Find 'Superstylin' by Groove Armada." In response to the initial multi-modal interaction, the ASR 620a may generate a preliminary interpretation that includes the words "Find" and "Superstylin" and the phrase "Groove Armada." The ASR 620a may then provide the preliminary interpretation to the conversational language processor 630a, which may determine that the word "Find" indicates that the most likely intent of the interaction includes a search, while the word "Superstylin" and the phrase "Groove Armada" provide criteria for the search. Furthermore, in response to determining the most likely intent of the interaction, the conversational language processor 630a may establish a music context for the interaction and attempt to resolve the search request. For example, the conversational language processor 630a may search one or more data repositories 636a that contain music information to identify music having a song title of "Superstylin" and an artist name of "Groove Armada," and the conversational language processor 630a may further cooperate with other devices in the hybrid processing environment to search for the song (e.g., in response to the local data repositories 636a not yielding adequate results, another device in the environment having a larger music data repository than the client device 610 may be invoked). The conversational language processor 630a may then receive appropriate results for the search and present the results to the user through the output device 615a (e.g., displaying information about the song, playing a sample audio clip of the song, displaying options to purchase the song, recommending similar songs or similar artists, etc.).

Continuing with the above example, a subsequent multi-modal interaction may include the utterance "Share this with my wife," "Dedicate it to Charlene," "That's the one, I want to pass along to some friends," or another suitable utterance that reflects a request to dedicate the content. Alternatively (or additionally), in response to the intent of the initial interaction including a request to search for content, the conversational language processor 630a may invoke the content dedication application 634a, which may present an option to dedicate the content through the output device 615a together with the results of the search. As such, the request to dedicate the content may also be provided in a non-speech input, such as a button press or touch screen

US 9,502,025 B2

29

selection of the option to dedicate the content. Accordingly, in response to detecting a content dedication request (e.g., through the ASR 620a and the conversational language processor 630a processing a suitable utterance, the input device 615a receiving a suitable non-speech input, or any suitable combination thereof), the content dedication application 634a may be invoked to process the content dedication request.

In one implementation, to process the content dedication request, the content dedication application 634a may capture a natural language utterance that contains the dedication. For example, the content dedication application 634a may provide a prompt through the output device 615a that instructs the user to provide the dedication utterance (e.g., a visual or audible prompt instructing the user to begin speaking, to speak after an audible beep, etc.). The user may then provide the dedication utterance through the voice-enabled input device 615a, and the dedication utterance may then be converted into an electronic signal that the content dedication application 634a captures for the dedication (e.g., “Dear Charlene, I was listening to this song and I thought of you. Enjoy!”). In addition, the content dedication application 634a may further prompt the user to provide any additional tags for the dedicated content (e.g., an image or a picture to be inserted as album art for the dedicated content, a natural language utterance that includes information to be inserted as voice-tags for the dedicated content, a non-speech input or data input that includes information to insert in tags for the dedicated content, etc.). The content dedication application 634a may then prompt the user to identify a recipient 690 of the dedication, wherein the user may provide any suitable multi-modal input that includes an e-mail address, a telephone number, an address book entry, or other information identifying the recipient 690 of the dedication.

In one implementation, the content dedication application 634a may then route the request, including the dedication utterance, the additional tags (if any), and the information identifying the recipient 690 to the content dedication system 665 through the messaging interface 650. For example, the dedication utterance may be converted into encoded audio that can be communicated through the messaging interface 650, whereby the content dedication system 665 can insert the encoded audio corresponding to the dedication utterance within the dedicated content and/or verbally annotate the dedicated content with the encoded audio. Alternatively (or additionally), the dedication utterance may be interpreted and parsed to transcribe one or more words or phrases from the dedication utterance, wherein the transcribed words or phrases may provide a textual annotation for the dedicated content (e.g., the textual annotation may be inserted within metadata for the dedicated content, such as an ID3 Comments tag). Similarly, any utterances to be inserted as voice-tags for the dedicated content may provide further verbal annotations for the dedicated content, or such utterances may be transcribed to provide further textual annotations for the dedicated content. In one implementation, verbal annotations, textual annotations, and other types of annotations may be created and associated with the dedicated content using techniques described in co-pending U.S. patent application Ser. No. 11/212,693, entitled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions,” filed Aug. 29, 2005, the contents of which are hereby incorporated by reference in their entirety.

In one implementation, in response to the content dedication system 665 receiving the content dedication request, the dedication utterance, and any additional tags from the

30

content dedication application 634a, the content dedication system 665 may invoke a similar content dedication application 634b on the voice-enabled server 640 to process the request. In particular, the content dedication application 634b on the voice-enabled server 640 may identify the content requested for dedication and initiate a transaction for the content requested for dedication. For example, in one implementation, if the content dedication application 634a on the voice-enabled client device 610 was able to identify the requested content, the content dedication request received at the content dedication system 665 may include an identification of the content to be dedicated. Alternatively, because the content dedication system 665 can cooperate in resolving the multi-modal interactions involved in the dedication request, the content dedication system 665 may use shared knowledge relating to the dedication request to identify the content to be dedicated. For example, the content dedication system 665 may invoke one or more local natural language processing components (e.g., ASR 620b, conversational language processor 630b or 630c, etc.), search one or more local data repositories 636b, interact with one or more content providers 680 over a network 670, pull data from a satellite radio system that played the requested content at the voice-enabled client device 610 or another device in the hybrid processing environment, or otherwise consult available resources that can be used to identify the content to be dedicated.

In one implementation, in response to identifying the content to be dedicated, the content dedication system 665 may communicate with a billing system 638 to identify one or more purchase options for the dedication request and process an appropriate transaction for the dedication request. In particular, the content dedication system 665 may generally support various different purchase options to provide users with flexibility in the manner of requesting content dedications, including a buy-to-own purchase option, a pay-to-play purchase option, a paid subscription purchase option, or other appropriate options. In one implementation, the purchase options may be modeled on techniques for providing natural language services and subscriptions described in U.S. patent application Ser. No. 10/452,147, entitled “Systems and Methods for Responding to Natural Language Speech Utterance,” which issued as U.S. Pat. No. 7,398,209 on Jul. 8, 2008, and co-pending U.S. patent application Ser. No. 10/618,633, entitled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance,” filed Jun. 15, 2003, the contents of which are hereby incorporated by reference in their entirety. For example, in the buy-to-own purchase option, the content dedication system 665 may purchase full rights to the content from an appropriate content provider 680, wherein the billing system 638 may then charge the user of the voice-enabled client device 610 a particular amount that encompasses the cost for purchasing the content from the content provider 680 plus a service charge for dedicating the content, tagging the dedicated content, delivering the dedicated content to the recipient 690, or any other appropriate services rendered for the content dedication request. The billing system 638 may charge the user in a similar manner under the pay-to-play purchase option, except that the rights purchased from the content provider 680 may be limited to a single play, such that the cost for purchasing the content from the content provider 680 may be somewhat less under the pay-to-play purchase option.

Under the paid subscription purchase option, however, the user may pay a periodic service charge to the content dedication system 665 that permits the user to make a

US 9,502,025 B2

31

predetermined number of content dedications or an unlimited number of content dedications in a subscription period, or the subscription purchase option may permit the user to make content dedications in any other suitable manner (e.g., different subscription levels having different content dedication options may be offered, such that the user may select a subscription level that meets the user's particular needs). For example, under the paid subscription purchase option, the billing system 638 may only charge the user any costs for obtaining the rights to the content, which may be purchased under either the buy-to-own option or the pay-to-play option, or alternatively the user may be charged nothing if the user already owns the dedication content. In another example, a first subscription level may cost a first amount to permit the user to make a particular number of content dedications in a subscription period, while a second subscription level may cost a higher amount to permit the user to make an unlimited number of content dedications in the subscription period, while still other subscription levels having different terms may be offered, as will be apparent.

Furthermore, in one implementation, a service provider associated with the content dedication system 665 may negotiate an agreement with the content provider 680 to determine the manner in which revenues for content transactions will be shared between the content dedication system 665 and the content provider 680. For example, the agreement may provide that the content provider 680 may keep all of the revenue for transactions that include purchasing content from the content provider 680 and that the service provider associated with the content dedication system 665 may recoup costs for such transactions from users. In another example, the agreement may provide that the content provider 680 and the service provider associated with the content dedication system 665 manage the revenue for the transactions that include purchasing content from the content provider 680, while the service provider associated with the content dedication system 665 may manage billing users for the natural language aspects of the content dedication service according to the techniques described in further detail above.

According to one aspect of the invention, FIG. 7 illustrates a flow diagram of an exemplary method for providing a natural language content dedication service. In particular, the method for providing a natural language content dedication service, as shown in FIG. 7, may be performed on a voice-enabled client device that can communicate with a content dedication system through a messaging interface, wherein the voice-enabled client device may communicate with the content dedication system in a similar manner as described above with reference to FIGS. 2 through 6. For example, the content dedication system may include a virtual router and a voice-enabled server that can service multi-modal natural language requests provided to the voice-enabled client device, and may further include a billing system for processing transactions relating to the content dedication service.

In one implementation, the method shown in FIG. 7 may be used to provide the natural language content dedication service to any suitable voice-enabled client device having a suitable combination of input and output devices that can receive natural language multi-modal interactions and provide responses to the natural language multi-modal interactions,

32

wherein the input and output devices may be further arranged to receive any other suitable type of input and provide any other suitable type of output. For example, in one implementation, the voice-enabled client device may comprise a mobile phone that includes a keypad input device, a touch screen input device, or other input mechanisms in addition to any input microphones or other suitable input devices that can receive voice signals, and may further include an output display device in addition to any output microphones or other suitable output devices that can output audible signals. Thus, a user of the voice-enabled client device may be listening to music, watching video, or otherwise interacting with content through the input and output devices, wherein an operation 710 may include the voice-enabled client device receiving a multi-modal natural language interaction to engage in a transaction to dedicate the music, video, or other content.

Thus, in one implementation, in response to the voice-enabled client device receiving the multi-modal interaction in operation 710 that includes a natural language utterance, the voice-enabled client device may invoke an Automatic Speech Recognizer (ASR) to generate one or more preliminary interpretations of the utterance. The ASR may then provide the preliminary interpretations to a conversational language processor, which may attempt to determine an intent for the multi-modal interaction. In one implementation, to determine the intent for the multi-modal interaction, the conversational language processor may determine a most likely context for the interaction from the preliminary interpretations of the utterance, any accompanying non-speech inputs in the multi-modal interaction that relate to the utterance, contexts associated with prior requests, short-term and long-term shared knowledge, or any other suitable information for interpreting the multi-modal interaction. Thus, in one implementation, an operation 720 may include the conversational language processor detecting a content dedication request in response to determining that the intent for the multi-modal interaction relates to a content dedication request. The conversational language processor may then invoke a content dedication application to resolve the content dedication request.

For example, in one implementation, the multi-modal interaction received in operation 720 may include the utterance "Buy that song and send it to Michael." In response to the initial multi-modal interaction, the ASR may generate a preliminary interpretation that includes words and/or phrases such as "Buy," "that song," "send it," and "to Michael." The ASR may then provide the preliminary interpretation to the conversational language processor, which may determine that the word and/or phrase combination of "Buy" and "send it" indicates that the most likely intent of the interaction includes a content dedication request, while the word and/or phrase combination of "that song" and "to Michael" provide criteria for the intended content and recipient for the dedication. Furthermore, in response to determining the most likely intent of the interaction, the conversational language processor may establish a device context, a dedication context, a content or music context, an address book context, or other suitable contexts in an attempt to resolve the request.

For example, the device context may enable the conversational language processor to retrieve data from the voice-enabled client device or another suitable device that provides the user's intended meaning for the phrase "that song" (e.g., the user may be referring to a song playing on the user's satellite radio device, such that the conversational language processor can identify the song that was playing

US 9,502,025 B2

33

when the device interaction was received in operation 710). Furthermore, the address book context may enable the conversational language processor to identify “Michael,” the intended recipient of the dedication request. The conversational language processor may then receive appropriate results for resolving the intent of the request and present the results to the user through the output device (e.g., displaying information requesting that the user confirm that the content and recipient was correctly identified, playing a sample audio clip of the song, displaying options to purchase the song, recommending similar songs or similar artists, etc.). Accordingly, in response to detecting the content dedication request in operation 720 and identifying the relevant criteria identifying the content to be dedicated and the intended recipient, the content dedication application may be invoked to process the content dedication request.

For example, in one implementation, processing the content dedication request may include the content dedication application capturing a natural language utterance that contains the dedication for the requested content in an operation 730. For example, the content dedication application may provide a prompt through the output device that instructs the user to provide the dedication utterance (e.g., a visual or audible prompt instructing the user to begin speaking, to speak after an audible beep, etc.). The user may then provide the dedication utterance through the voice-enabled input device, and the dedication utterance may then be converted into an electronic signal that the content dedication application captures in operation 730. In addition, an operation 740 may include the content dedication application further prompting the user to provide any additional tags for the dedicated content (e.g., an image or a picture to be inserted as album art for the dedicated content, one or more natural language utterances to insert as voice-tags for the dedicated content, one or more natural language utterances to be transcribed into text that can be inserted in tags for the dedicated content, a non-speech input or data input that include information to insert in tags for the dedicated content, etc.).

Thus, in response to determining that the user has provided additional information to insert in tags for the dedication content in operation 740, the content dedication application may then capture the tags in an operation 750. For example, operation 750 may include the content dedication application capturing an image or picture that the user identifies for album art to be inserted in the dedicated content, capturing any natural language utterances to insert as voice-tags for the dedicated content, communicating with the ASR and/or conversational language processor to transcribe any natural language utterances to be inserted as text within tags for the dedicated content, capturing any non-speech inputs or data inputs that include information to insert as text within the tags for the dedicated content, or otherwise capturing information that relates to the additional tags.

In one implementation, the content dedication application may then process the content dedication request in an operation 760, which may include prompting the user to identify the recipient of the dedication. For example, the content dedication application may request information identifying the recipient of the dedication in response to the user not having already identified the recipient, in response to the user identifying the recipient in a manner that includes ambiguity or other criteria that cannot be resolved without further information, to distinguish among different contact information known for the recipient, or in response to other appropriate circumstances. Thus, the user may provide any

34

suitable multi-modal input that includes an e-mail address, a telephone number, an address book entry, or other criteria that can be used to uniquely identify the information for contacting the recipient of the dedication. Furthermore, in one implementation, processing the content dedication request in operation 760 may further include the content dedication application routing the request, including the dedication utterance, the additional tags (if any), and the information for contacting the recipient to the content dedication system through the messaging interface, wherein the content dedication system may then process a transaction for the content dedication request, as will be described in greater detail below.

According to one aspect of the invention, FIG. 8 illustrates a flow diagram of an exemplary method for providing a natural language content dedication service. In particular, the method for providing a natural language content dedication service, as shown in FIG. 8, may include an operation 810 in which a content dedication system may receive a natural language content dedication request through a messaging interface. For example, in one implementation, the content dedication system may receive the natural language content dedication request from a content dedication application that executes on a voice-enabled client device. In one implementation, the natural language content dedication request received in operation 810 may generally include a natural language dedication utterance, information to be inserted within one or more tags for content to be dedicated, and information identifying a recipient of the dedicated content.

For example, in one implementation, the dedication utterance received in operation 810 may include encoded audio that the content dedication system can insert within the dedicated content or use to verbally annotate the dedicated content. Alternatively (or additionally), the content dedication system may interpret and parse the dedication utterance to transcribe one or more words or phrases from the dedication utterance, wherein the transcribed words or phrases may provide a textual annotation for the dedicated content (e.g., the textual annotation may be inserted within metadata for the dedicated content, such as an ID3 Comments tag). Similarly, any utterances to be inserted as voice-tags for the dedicated content may provide further verbal annotations for the dedicated content, or such utterances may be transcribed to provide further textual annotations for the dedicated content. In one implementation, verbal annotations, textual annotations, and other types of annotations may be created and associated with the dedicated content using techniques described in co-pending U.S. patent application Ser. No. 11/212,693, entitled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions,” filed Aug. 29, 2005, the contents of which are hereby incorporated by reference in their entirety.

In one implementation, in response to the content dedication system receiving the content dedication request, the dedication utterance, and any additional tags from the content dedication application in operation 810, the content dedication system may invoke a content dedication application on a voice-enabled server to process the content dedication request. In particular, an operation 820 may include the content dedication application on the voice-enabled server identifying the content requested for dedication. In one implementation, the request received from the content dedication application in operation 810 may include information identifying the requested content and/or information that the content dedication system can use to identify the requested content. For example, the request may include

US 9,502,025 B2

35

a multi-modal natural language input that includes a natural language utterance and/or a non-voice input, wherein the content dedication system may use shared knowledge relating to the dedication request to identify the content to be dedicated.

Thus, operation **820** may include the content dedication system invoking one or more natural language processing components (e.g., an ASR, a conversational language processor, etc.), searching one or more data repositories, interacting with one or more content providers, pulling data from a satellite radio system that played the requested content at the voice-enabled client device or another device in a hybrid processing environment, or otherwise consulting available resources that can be used to identify the content to be dedicated. For example, in one implementation, the hybrid processing environment may include a device having an application that can identify played content (e.g., a Shazam® listening device that a user can hold near a speaker to identify content playing through the speaker). Thus, operation **820** may generally include the content dedication system communicating with any suitable device, system, application, or other resource to identify the content requested for dedication.

In one implementation, in response to identifying the content to be dedicated, an operation **830** may include the content dedication system identifying one or more purchase options for the dedication request. In one implementation, operation **830** may include the content dedication system communicating with a billing system that supports various purchase options to provide users with flexibility in the manner of requesting content dedications. For example, the billing system may support content dedication purchase options that include a buy-to-own purchase option, a pay-to-play purchase option, a paid subscription purchase option, or other appropriate options. In response to identifying the purchase option for the content dedication request, an operation **840** may include the content dedication system processing a transaction for the content identified in operation **820**.

For example, in response to operation **830** indicating that the user has selected the buy-to-own purchase option, the transaction processed in operation **840** may include purchasing full rights to the requested content from an appropriate content provider, wherein the billing system may then charge the user of the voice-enabled client device an appropriate amount that includes costs for purchasing the content from the content provider, adding the natural language utterance dedicating the content, tagging the dedicated content, delivering the dedicated content to the recipient, or any other appropriate services rendered for the content dedication request. Alternatively, in response to operation **830** indicating that the user has selected the pay-to-play purchase option, the transaction processed in operation **840** may include purchasing rights from the content provider that permits the purchased content to be played a predetermined number of times, wherein the billing system may charge the user in a similar manner as described above, except that the cost for purchasing limited rights to the content under the pay-to-play purchase option may be somewhat less than the cost for purchasing ownership rights to the content, as in the buy-to-own purchase option.

In another alternative implementation, in response to determining that the requested content dedication includes a selection of the paid subscription purchase option, the user may pay a periodic service charge to the content dedication system that permits the user to make a predetermined number of content dedications in a subscription period, an

36

unlimited number of content dedications in the subscription period, or otherwise make content dedications in accordance with terms defined in a subscription (e.g., different subscription levels having different content dedication options may be offered, such as a subscription level that only permits utterance dedications, another subscription level that further permits interpreting and parsing utterance dedications, etc.). Thus, the content transaction may be processed in operation **840** according to the purchase options identified in operation **830**, and the content dedication system may then further process the dedication request to customize the dedicated content according to criteria provided in the request previously received in operation **810**.

For example, in one implementation, an operation **850** may include the content dedication system inserting the natural language dedication utterance into the dedicated content, verbally annotating the dedicated content with the dedication utterance, or otherwise associating the dedicated content with the dedication utterance. Furthermore, an operation **860** may include the content dedication system determining whether any additional tags have been specified for the dedicated content. For example, as noted above, the dedication request may identify an image or picture to insert as album art for the dedicated content, one or more natural language utterances, non-voice, and/or data inputs to be transcribed into text that can be inserted in tags for the dedicated content, or any other suitable information that can be inserted within or associated with metadata for the dedicated content. Thus, in response to determining that information for any additional tags has been provided, an operation **870** may include inserting such additional tags into the dedicated content.

In one implementation, in response to having purchased the content requested for dedication, associating the dedication utterance with the dedicated content, and associating the additional tags (if any) with the dedicated content, the content dedication system may send a content dedication message to the recipient of the dedication in an operation **880**. For example, the content dedication message may include a Short Message Service (SMS) text message, an electronic mail message, an automated telephone call managed by a text-to-speech engine, or any other appropriate message that can be appropriately delivered to the recipient (e.g., the message may include a link that the recipient can select to stream, download, or otherwise access the dedicated content, the dedication utterance, etc.). Thus, the content dedication message may generally notify the recipient that content has been dedicated to the recipient and provide various mechanisms for the recipient to access the content dedication, as will be apparent.

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 cer-

US 9,502,025 B2

37

tain 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.

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 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.

What is claimed is:

1. A system for providing a natural language content dedication service, comprising:

one or more processors; and
 one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to:
 receive a first utterance that includes a natural language utterance;
 determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance;
 provide the one or more words or phrases as an input to a conversational language processor;
 interpret the first utterance, at the conversational language processor, based on the one or more words or phrases;
 identify, based on the interpretation of the first utterance, content to dedicate to a recipient;
 initiate, based on the identified content, a dedication to the recipient;
 receive a second utterance to be associated with the dedication;
 determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance;
 provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and
 send information to enable the recipient to access the content and the second utterance.

2. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an image as one or more tags within metadata of the content.

3. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to annotate the content with the second utterance.

4. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to purchase the content from a content provider to enable the recipient to access the content.

38

5. The system of claim 4, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to:

identify, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and
 bill a user associated with the first utterance based on the purchase option.

6. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: send, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and

send, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.

7. The system of claim 1, wherein the content includes one or more of a song or a video.

8. A method for providing a natural language content dedication service, the method being implemented in a computer system having one or more processors programmed with computer program instructions that, when executed by the one or more processors, cause the computer system to perform the method, the method comprising:

receiving, by one or more processors, a first utterance that includes a natural language utterance;
 determining, by the one or more processors, one or more words or phrases of the first utterance based on processing of the first utterance by a speech recognition engine;
 providing, by the one or more processors, the one or more words or phrases as an input to a conversational language processor;
 interpreting, by the one or more processors, the first utterance, at the conversational language processor, based on the one or more words or phrases;
 identifying, by the one or more processors, content to dedicate to a recipient based on the interpretation of the first utterance;
 initiating, by the one or more processors, a dedication to the recipient based on the identified content;
 receiving, by the one or more processors, a second utterance to be associated with the dedication;
 determining, by the one or more processors, one or more words or phrases of the second utterance based on a processing of the second utterance by the speech recognition engine;
 providing, by the one or more processors, the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and
 sending, by the one or more processors, information to enable the recipient to access the content and the second utterance.

9. The method of claim 8, further comprising providing an image as one or more tags within metadata of the content.

10. The method of claim 8, further comprising annotating the content with the second utterance.

11. The method of claim 8, further comprising purchasing the content from a content provider to enable the recipient to access the content.

12. The method of claim 11, further comprising:
 identifying, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and

US 9,502,025 B2

39

billing a user associated with the first utterance based on the purchase option.

13. The method of claim 8, further comprising: sending, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and 5 sending, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.

14. The method of claim 8, wherein the content includes one or more of a song or a video.

15. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to:

determine that a user is interacting with a media item; and 15 identify, based on the natural language processing and the interaction of the user with the media item, the media item as the content to dedicate to the recipient.

16. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to:

determine a media item that a user is playing; 20

40

determine that the user provided the first utterance during the playing of the media item by the user; and identify, based on the natural language processing and the determination that the user provided the first utterance during the playing of the media item, the media item as the content to dedicate to the recipient.

17. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an electronic link associated with the content as one or more tags within metadata of the content.

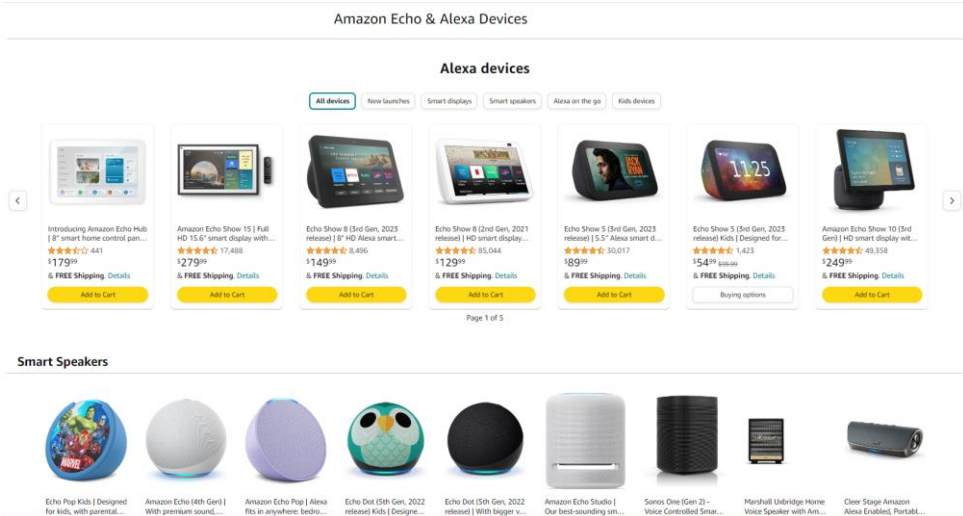
18. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide a voice tag as one or more tags within metadata of the content.

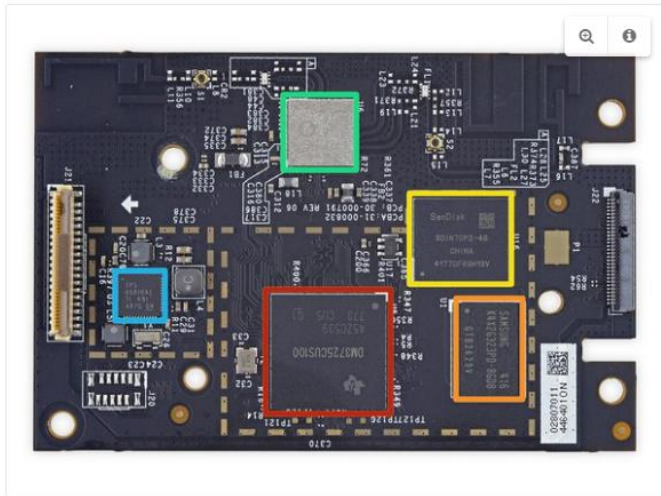
19. The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide the second utterance within the content.

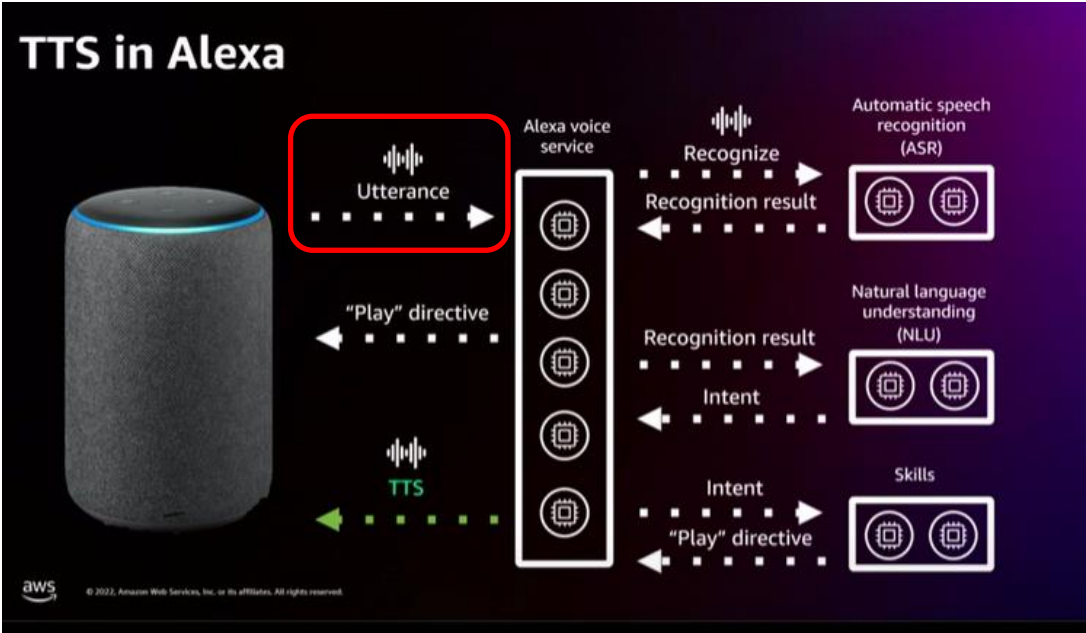
* * * * *

EXHIBIT F

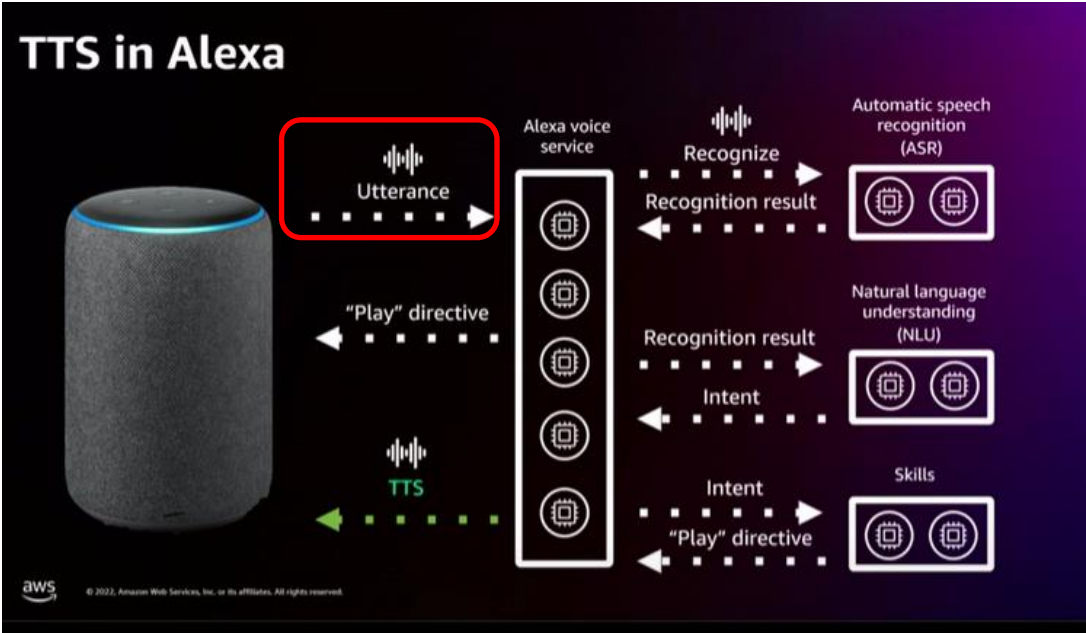
Exhibit F: Exemplary Infringement Claim Chart for U.S. Patent 10,297,249 Claim 1

Claim	Exemplary Disclosures
<p>[1.pre] A computer-implemented method of facilitating natural language system responses using short-term knowledge generated based on one or more prior multi-modal device interactions, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:</p>	<p>Amazon's Alexa is programmed to perform and have performed methods for facilitating natural language system responses using short-term knowledge generated based on one or more prior multi-modal device interactions, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method.</p> <p>Amazon's Alexa includes, for example, Echo, Echo Dot, Echo Show, FireTV, Alexa apps, and Alexa cloud.</p>  <p>See https://www.amazon.com/b?&node=9818047011&ref=MARS_NAV_desktop_plswitcher</p> <p>Alexa generates short-term knowledge based on multi-modal device interactions. For example, interactions with the touchscreen or the Echo Show or physical button presses for a FireTV and the resulting visual displays impact subsequent interpretation of a user's utterance.</p>

Claim	Exemplary Disclosures
	<p>Additionally, Amazon's Alexa that has one or more physical processors. See https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/</p> <p>Additionally, Echo devices have one or more processors. For example, the Texas Instruments DM3725CUS100 Digital Media Processor as depicted below with red annotations:</p>  <p>Sam Goldheart, Amazon Echo Teardown, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953</p>
<p>[1.a] receiving, by the computer system during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;</p>	<p>Alexa is programmed to perform and have performed methods for receiving, by the computer system during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance.</p> <p>Users can have multiple Alexa devices. For example, a user can have a first input device, an Echo show. An Echo show can receive a voice input such as “Alexa, search for [TV show],” which is a natural language utterance.</p>

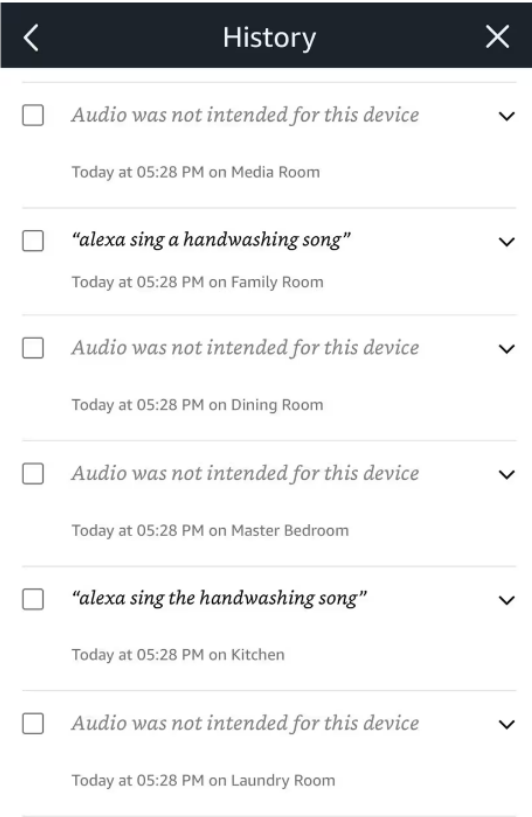
Claim	Exemplary Disclosures
	 <p>The diagram, titled "TTS in Alexa", illustrates the process of handling a user's voice command. On the left, an Alexa device is shown. A red box highlights the "Utterance" being sent from the device to the "Alexa voice service" (represented by a vertical stack of four server icons). From the Alexa voice service, the process flows through three main components: "Automatic speech recognition (ASR)", "Natural language understanding (NLU)", and "Skills". Each component sends a "Recognition result" back to the Alexa voice service. The "Skills" component also sends an "Intent" back. Finally, the Alexa voice service sends a "Play" directive back to the Alexa device, which then uses "TTS" (Text-to-Speech) to generate a response. The AWS logo and copyright notice "© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved." are visible at the bottom left of the diagram.</p> <p>See https://www.youtube.com/watch?v=gRAfp--sZ_Y at 10:07.</p>

Claim	Exemplary Disclosures
	<p>Digital Services and Device Support › Alexa Devices Help › Support for Echo Show ›</p> <h2>Watch Video Streaming Services on Echo Show</h2> <p>Use your voice to play movies, videos, or TV shows from Prime Video and supported streaming services. These steps also work with Echo Spot.</p> <hr/> <ul style="list-style-type: none"> • "Show me my video library." (Prime Video only) • "Show me my Watchlist." • "Open Fire TV" (Echo Show 15 only) • "Show me [title]." • "Search for [TV show]." • "Show me movies with [actor]." • "Show me [genre] movies." • "Open [video provider]." <p>If you have an Echo Show 15, you can also use commands like these to watch video or live content with Fire TV.</p> <p>Related Help Topics</p> <p>What is the All-New Fire TV Experience?</p> <p>Link Video Streaming Services on Echo Show</p> <p>Pair a Fire TV Remote With Your Echo Show 15</p> <p>See https://www.amazon.com/gp/help/customer/display.html?nodeId=GWMLYY6RFBGZDA29</p>
[1.b] receiving, by the computer system, a second voice input comprising the first natural language utterance via a second input device;	Alexa is programmed to perform and have performed methods of receiving, by the computer system, a second voice input comprising the first natural language utterance via a second input device.

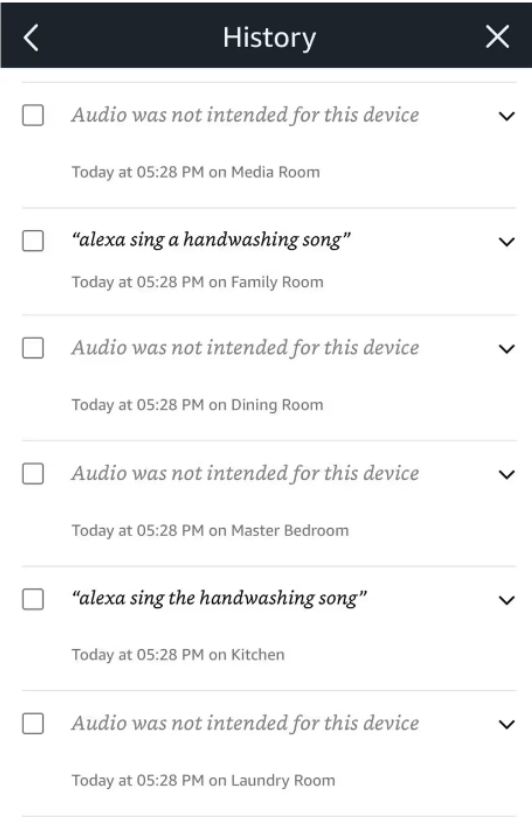
Claim	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>See https://www.youtube.com/watch?v=gRAfp--sZ_Y at 10:07.</p> <p>Users can have multiple Alexa devices. For example, a user can have a second input device, e.g., a Fire TV, in proximity to another Alexa device (e.g., Echo). A Fire TV can also receive voice input, for example via its remote controller.</p>

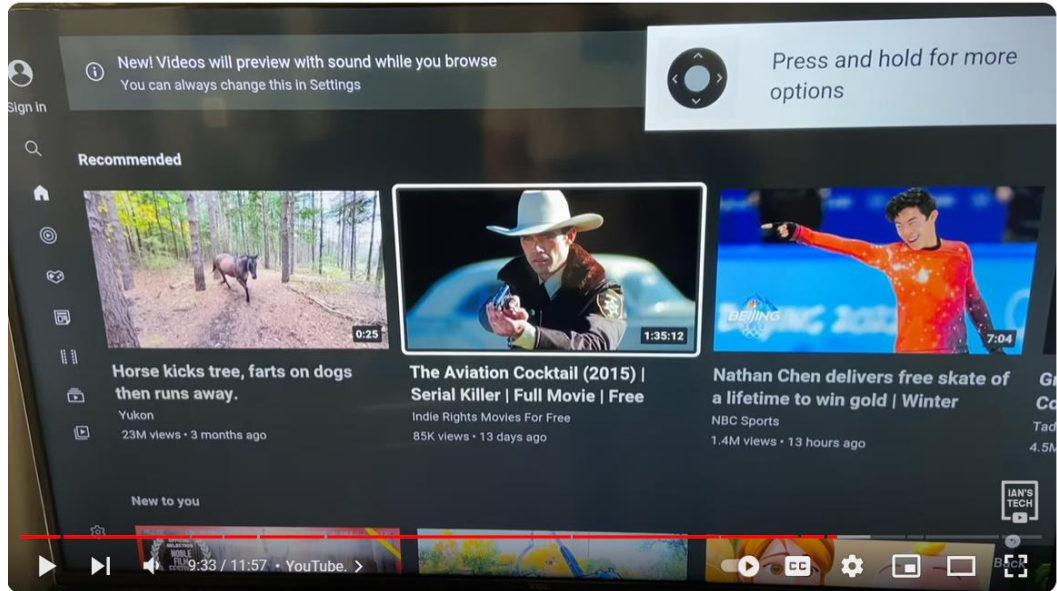
Claim	Exemplary Disclosures										
	<p>Control Your Fire TV</p> <p>Once you've completed the linking process, you can use your Alexa-enabled devices to control the following features on your Fire TV.</p> <table border="1"> <thead> <tr> <th>To do this on your Fire TV...</th><th>Say this to your Alexa device...</th></tr> </thead> <tbody> <tr> <td> Play a movie or TV show Alexa can play content from Prime Video and other supported apps installed on your Fire TV. </td><td> "Watch [title]." "Play [title]." "Play [title] on [app]." </td></tr> <tr> <td> Search for movies and TV shows Alexa can find movies and TV shows from Prime Video and supported installed apps. </td><td> "Search for [title]." "Find [title / genre]." "Show me [tv shows / movies] with [actor]." "Show me new [tv shows / movies]." "Search for free [tv shows / movies]." </td></tr> <tr> <td> Play music Alexa can play music from your connected audio providers. </td><td> "Play [song / album / artist]." "Play the latest episode of my podcast." "Play relaxing music." </td></tr> <tr> <td> Navigate search results Alexa can navigate through search results and select items by voice. </td><td> "Scroll [direction]." "Select #3." "Play #6." "Go back." </td></tr> </tbody> </table> <p>See https://www.amazon.com/gp/help/customer/display.html?nodeId=202174250</p> <p>As a non-limiting example, a FireTV receives the same voice input as an Echo show receives, which is "Alexa, search for [TV show]," the same natural language utterance.</p>	To do this on your Fire TV...	Say this to your Alexa device...	Play a movie or TV show Alexa can play content from Prime Video and other supported apps installed on your Fire TV.	"Watch [title]." "Play [title]." "Play [title] on [app]." 	Search for movies and TV shows Alexa can find movies and TV shows from Prime Video and supported installed apps.	"Search for [title]." "Find [title / genre]." "Show me [tv shows / movies] with [actor]." "Show me new [tv shows / movies]." "Search for free [tv shows / movies]." 	Play music Alexa can play music from your connected audio providers.	"Play [song / album / artist]." "Play the latest episode of my podcast." "Play relaxing music." 	Navigate search results Alexa can navigate through search results and select items by voice.	"Scroll [direction]." "Select #3." "Play #6." "Go back."
To do this on your Fire TV...	Say this to your Alexa device...										
Play a movie or TV show Alexa can play content from Prime Video and other supported apps installed on your Fire TV.	"Watch [title]." "Play [title]." "Play [title] on [app]." 										
Search for movies and TV shows Alexa can find movies and TV shows from Prime Video and supported installed apps.	"Search for [title]." "Find [title / genre]." "Show me [tv shows / movies] with [actor]." "Show me new [tv shows / movies]." "Search for free [tv shows / movies]." 										
Play music Alexa can play music from your connected audio providers.	"Play [song / album / artist]." "Play the latest episode of my podcast." "Play relaxing music." 										
Navigate search results Alexa can navigate through search results and select items by voice.	"Scroll [direction]." "Select #3." "Play #6." "Go back."										
[1.c] comparing, by the computer system, the first voice input with the second voice input;	<p>Alexa is programmed to perform and have performed methods of comparing, by the computer system, comparing, by the computer system, the first voice input with the second voice input.</p> <p>When different Accused Products are within proximity of one another, the Alexa system compares the first voice input, such as an Echo show, with the second voice input, such as a FireTV by comparing at least the wake words and the distance.</p>										

Claim	Exemplary Disclosures
	<p>“If that's not the issue then it may be worth changing the Alexa wake word on some of your devices. That way, some of your devices are not overhearing the Alexa wake word. Here's how you can change the wake word:</p> <p>Open the Alexa app. Open Devices. Select All Devices and then select your device. If your device has active Reminders or Routines, you might have to select settings to reach the Device Settings page. Scroll under General and select Wake Word. Select a wake word from the list, and then select OK.”</p> <p>See https://www.amazonforum.com/s/question/0D54P0000898f9ASAQ/how-to-stop-wrong-alexa-device-from-responding</p> <p>“Another way you could separate many Alexa devices is by changing their wake word.</p> <p>“There are millions of customers with more than one Alexa-enabled device in their home, and for these customers one of the most important features we’ve built is Echo Spatial Perception (ESP). The ESP feature provides a frustration-free experience by identifying the best device to respond when interacting with Alexa in a multi-device environment. The technology determines which device is closest to the customer, and helps Alexa respond through that device.” https://developer.amazon.com/en-US/blogs/alexa/post/042be85c-5a62-4c55-a18d-d7a82cf394df/esp-moves-to-the-cloud-for-alexa-enabled-device</p>

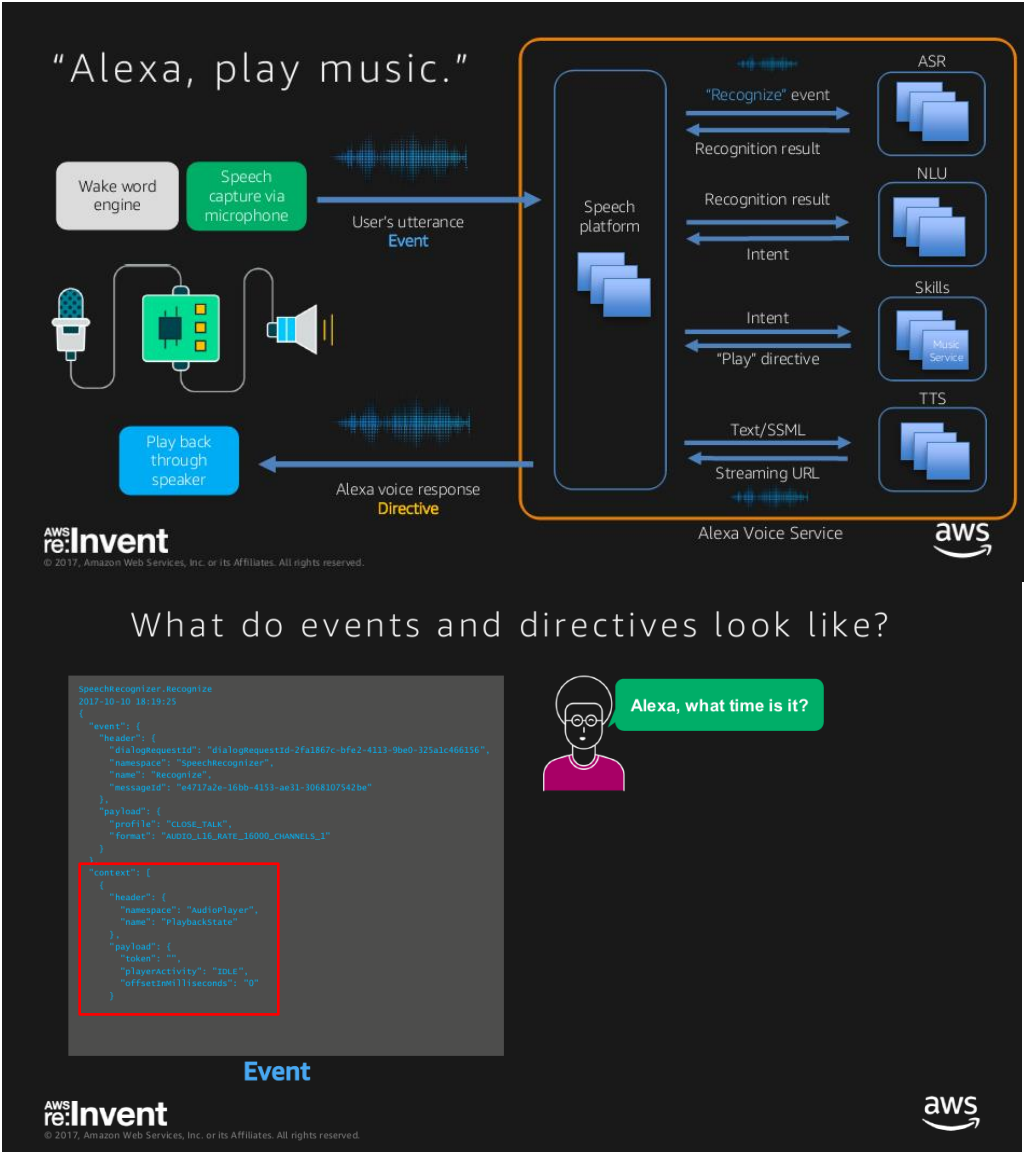
Claim	Exemplary Disclosures
	 <p>History</p> <ul style="list-style-type: none"> ☐ Audio was not intended for this device ✓ Today at 05:28 PM on Media Room ☐ "alexa sing a handwashing song" ✓ Today at 05:28 PM on Family Room ☐ Audio was not intended for this device ✓ Today at 05:28 PM on Dining Room ☐ Audio was not intended for this device ✓ Today at 05:28 PM on Master Bedroom ☐ "alexa sing the handwashing song" ✓ Today at 05:28 PM on Kitchen ☐ Audio was not intended for this device ✓ Today at 05:28 PM on Laundry Room <p>https://amazonforum.my.site.com/s/question/0D54P00007WnYgtSAF/multiple-alexa-devices-responding-why-is-echo-spatial-perception-not-working</p>
[1.d] filtering, by the computer system, sound from the first voice input and the second voice input based on the comparison;	<p>Alexa is programmed to perform and have performed methods of filtering, by the computer system, sound from the first voice input and the second voice input based on the comparison.</p> <p>While multiple Alexa devices may be within earshot of the end-user, the Echo device within the closest proximity to the end-user presumes that it was the intended recipient of the utterance and responds accordingly.</p>

Claim	Exemplary Disclosures
	<p>“If that's not the issue then it may be worth changing the Alexa wake word on some of your devices. That way, some of your devices are not overhearing the Alexa wake word. Here's how you can change the wake word:</p> <p>Open the Alexa app. Open Devices. Select All Devices and then select your device. If your device has active Reminders or Routines, you might have to select settings to reach the Device Settings page. Scroll under General and select Wake Word. Select a wake word from the list, and then select OK.”</p> <p>See https://www.amazonforum.com/s/question/0D54P0000898f9ASAQ/how-to-stop-wrong-alexa-device-from-responding</p> <p>Alexa wakes the right device by filtering sound from the first voice input and the second voice input based on the comparison of at least the distance of the Alexa devices.</p> <p>“There are millions of customers with more than one Alexa-enabled device in their home, and for these customers one of the most important features we’ve built is Echo Spatial Perception (ESP). The ESP feature provides a frustration-free experience by identifying the best device to respond when interacting with Alexa in a multi-device environment. The technology determines which device is closest to the customer, and helps Alexa respond through that device.” https://developer.amazon.com/en-US/blogs/alexa/post/042be85c-5a62-4c55-a18d-d7a82cf394df/esp-moves-to-the-cloud-for-alexa-enabled-device</p>

Claim	Exemplary Disclosures
	 <p>https://amazonforum.my.site.com/s/question/0D54P00007WnYgtSAF/multiple-alexa-devices-responding-why-is-echo-spatial-perception-not-working</p>
<p>[1.e] obtaining, by the computer system during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input;</p>	<p>Alexa is programmed to perform and have performed methods of obtaining, by the computer system during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input.</p> <p>As a non-limiting example, if a user wants to play “the Aviation Cocktail” from Youtube, the user can say “Play this” (a voice input) while using the FireTV remote to identify the correct video to play from a list of recommendation (a non-voice input).</p>

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=wROAsbo1DQ8</p>
<p>[1.f] generating, by the computer system, the short-term knowledge based on at least the first voice input and the first non-voice input;</p>	<p>Alexa is programmed to perform and have performed methods of generating, by the computer system, the short-term knowledge based on at least the first voice input and the first non-voice input.</p> <p>“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p>“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p>

Claim	Exemplary Disclosures										
	<p>https://www.amazon.science/latest-news/the-engineering-behind-alexa-contextual-speech-recognition</p> <p>Contextual carryover</p> <p>You don't need to provide annotated dialogs that demonstrate contextual carryover. The dialog simulator supports this feature by introducing turns with pronouns (for example, "Please purchase it") when generating dialog variants from annotated dialogs. The carryover occurs at a later runtime stage, argument filling, when Alexa Conversations considers all slots that the user and Alexa mention across the entire dialog for filling API arguments.</p> <p>Example</p> <p>You provide a dialog to purchase a movie. The dialog simulator introduces a turn with a pronoun.</p> <table border="1"> <thead> <tr> <th data-bbox="821 708 1297 756">Dialog you provide</th><th data-bbox="1297 708 1814 756">Example variant from the dialog simulator</th></tr> </thead> <tbody> <tr> <td data-bbox="821 769 1297 824">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td><td data-bbox="1297 769 1814 824">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td></tr> <tr> <td data-bbox="821 828 1297 852">Alexa: How about Snatch?</td><td data-bbox="1297 828 1814 852">Alexa: How about Snatch?</td></tr> <tr> <td data-bbox="821 855 1297 911">User: Please help me purchase Snatch. (Invoke API ReserveMovie.)</td><td data-bbox="1297 855 1814 911">User: Please purchase it. (Invoke API ReserveMovie.)</td></tr> <tr> <td data-bbox="821 914 1297 938">Alexa: The movie Snatch is reserved for you.</td><td data-bbox="1297 914 1814 938">Alexa: The movie Snatch is reserved for you.</td></tr> </tbody> </table> <p>Context memory</p> <p><i>Context memory</i> is short-term memory that tracks user utterances and models results such as entities (that is, a slot that the user mentioned or a return value from a previous API), predicted actions, and Alexa responses. Context memory maintains dialog context for each session and is constantly updated throughout the session.</p> <p>See https://developer.amazon.com/en-US/docs/alexa/conversations/how-alexa-conversations-works.html#contextual-carryover</p>	Dialog you provide	Example variant from the dialog simulator	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	Alexa: How about Snatch?	Alexa: How about Snatch?	User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)	Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.
Dialog you provide	Example variant from the dialog simulator										
User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)										
Alexa: How about Snatch?	Alexa: How about Snatch?										
User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)										
Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.										

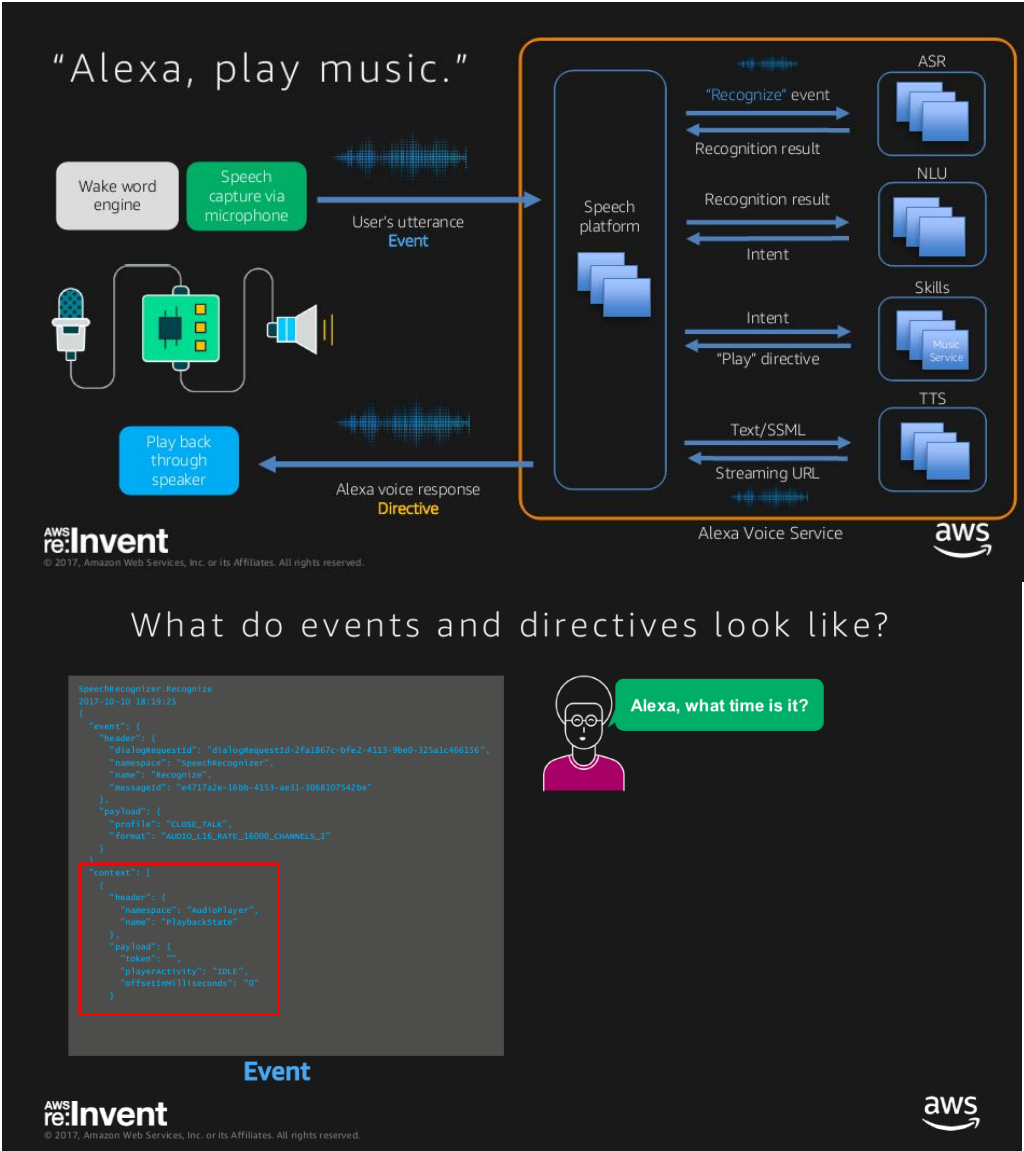
Claim	Exemplary Disclosures
	 <p>The diagram illustrates the Alexa Voice Service architecture. A user's utterance is captured via a microphone and processed by a wake word engine and speech capture. The utterance is then sent to the Speech platform, which interacts with the ASR (Automatic Speech Recognition) service to recognize the event. The platform then sends the recognition result to the NLU (Natural Language Understanding) service to determine the intent. Based on the intent, the platform sends a directive (e.g., "Play" directive) to the Skills service (e.g., Music Service). The Skills service then sends a Text/SSML response to the TTS (Text-to-Speech) service, which generates a streaming URL. The Speech platform then plays back the response through a speaker.</p> <p>Below the diagram, the text "What do events and directives look like?" is followed by an example of an event JSON payload:</p> <pre> { "event": { "header": { "dialogRequestId": "dialogRequestId-2fa187c-bfe2-4113-9be0-325a1c466156", "namespace": "SpeechRecognizer", "name": "Recognize", "messageId": "e4717a2e-16bb-4153-ae31-3068107542be" }, "payload": { "profile": "CLOSE_TALK", "format": "AUDIO_L16_RATE_16000_CHANNELS_1" } }, "context": { "header": { "namespace": "AudioPlayer", "name": "Playbackstate" }, "payload": { "token": "...", "playerActivity": "SILENCE", "offsetInMilliseconds": "0" } } } </pre> <p>The JSON payload is labeled "Event" and is associated with the user's utterance "Alexa, what time is it?".</p>

AWS re: Invent, Integrate Alexa into Your Product, Using the Alexa Voice Service Device SDK.

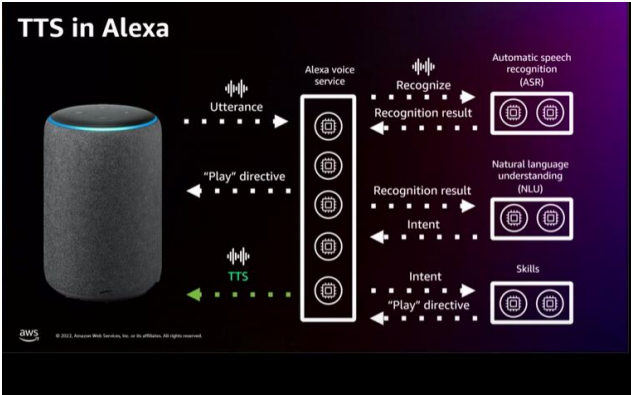
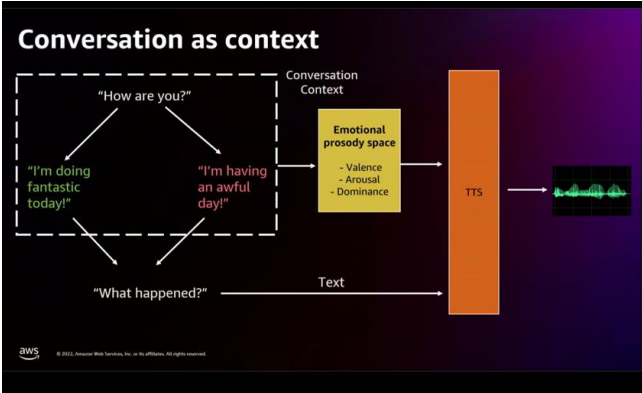
Claim	Exemplary Disclosures
	<div data-bbox="686 211 1312 604"> <p>TTS in Alexa</p> <p>The diagram illustrates the TTS process in Alexa. It starts with an 'Utterance' (represented by a microphone icon) being sent to the 'Alexa voice service'. This service then branches into three paths: 'Automatic speech recognition (ASR)', 'Natural language understanding (NLU)', and 'Skills'. Each path receives a 'Recognition result' or 'Intent' and sends a 'Play' directive back to the Alexa device, which then uses 'TTS' (Text-to-Speech) to produce the final output.</p> </div> <div data-bbox="686 638 1323 1027"> <p>Conversation as context</p> <p>This diagram shows how conversation context is used for TTS. It features a 'Conversation Context' box containing three utterances: 'How are you?', 'I'm doing fantastic today!', and 'I'm having an awful day!'. These utterances are processed through an 'Emotional prosody space' (which includes 'Valence', 'Arousal', and 'Dominance') and then through a 'TTS' block to produce the final audio output. A 'Text' input 'What happened?' is also shown entering the TTS block.</p> </div> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y</p>
<p>[1.g] determining, by the computer system, based on the short-term knowledge, a first context for the first natural language utterance;</p>	<p>Alexa is programmed to perform and have performed methods of determining, by the computer system, based on the short-term knowledge, a first context for the first natural language utterance.</p> <p>Alexa determines the context which video to play when the user utters “play this” or “play [number]” based on the short-term knowledge.</p>

Claim	Exemplary Disclosures
	<p data-bbox="722 248 1604 272">Digital Services and Device Support › Alexa Features Help › Video with Alexa ›</p> <h2 data-bbox="722 285 1583 334">Play Videos with Alexa Using Your Voice</h2> <p data-bbox="722 376 1623 443">Use your voice to play movies, videos, or TV shows from Prime Video and supported streaming services on Alexa devices and Fire TV devices.</p> <p data-bbox="699 483 861 508">Video Search*</p> <p data-bbox="699 534 1703 592">After searching for video content, say "Select [video]" or "Play [number]" to choose a video from the results.</p> <p data-bbox="699 630 1136 654">Actions on Detail Page* (Fire TV Only)</p> <ul data-bbox="722 690 1041 784" style="list-style-type: none"> • "Play this." • "Add this to my Watchlist." • "Play the trailer." <p data-bbox="699 813 1713 872">Note: * Not all features are supported from a paired Echo Show or for a paired Fire TV Cube (2nd Generation).</p> <p data-bbox="688 907 1927 938">See https://www.amazon.com/gp/help/customer/display.html?nodeId=G3YFGTLKXC2YWLQ</p> <p data-bbox="688 1013 1934 1149">“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p data-bbox="688 1190 1934 1398">“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p>

Claim	Exemplary Disclosures										
	<p>https://www.amazon.science/latest-news/the-engineering-behind-alexa-contextual-speech-recognition</p> <p>Contextual carryover</p> <p>You don't need to provide annotated dialogs that demonstrate contextual carryover. The dialog simulator supports this feature by introducing turns with pronouns (for example, "Please purchase it") when generating dialog variants from annotated dialogs. The carryover occurs at a later runtime stage, argument filling, when Alexa Conversations considers all slots that the user and Alexa mention across the entire dialog for filling API arguments.</p> <p>Example</p> <p>You provide a dialog to purchase a movie. The dialog simulator introduces a turn with a pronoun.</p> <table border="1"> <thead> <tr> <th data-bbox="821 708 1297 756">Dialog you provide</th><th data-bbox="1297 708 1814 756">Example variant from the dialog simulator</th></tr> </thead> <tbody> <tr> <td data-bbox="821 769 1297 824">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td><td data-bbox="1297 769 1814 824">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td></tr> <tr> <td data-bbox="821 829 1297 852">Alexa: How about Snatch?</td><td data-bbox="1297 829 1814 852">Alexa: How about Snatch?</td></tr> <tr> <td data-bbox="821 857 1297 912">User: Please help me purchase Snatch. (Invoke API ReserveMovie.)</td><td data-bbox="1297 857 1814 912">User: Please purchase it. (Invoke API ReserveMovie.)</td></tr> <tr> <td data-bbox="821 917 1297 940">Alexa: The movie Snatch is reserved for you.</td><td data-bbox="1297 917 1814 940">Alexa: The movie Snatch is reserved for you.</td></tr> </tbody> </table> <p>Context memory</p> <p><i>Context memory</i> is short-term memory that tracks user utterances and models results such as entities (that is, a slot that the user mentioned or a return value from a previous API), predicted actions, and Alexa responses. Context memory maintains dialog context for each session and is constantly updated throughout the session.</p> <p>See https://developer.amazon.com/en-US/docs/alexa/conversations/how-alexa-conversations-works.html#contextual-carryover</p>	Dialog you provide	Example variant from the dialog simulator	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	Alexa: How about Snatch?	Alexa: How about Snatch?	User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)	Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.
Dialog you provide	Example variant from the dialog simulator										
User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)										
Alexa: How about Snatch?	Alexa: How about Snatch?										
User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)										
Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.										

Claim	Exemplary Disclosures
	 <p>The diagram illustrates the Alexa Voice Service architecture. It shows a user's utterance being captured via a microphone and processed by a wake word engine and speech capture. The utterance is then sent to the Speech platform, which interacts with the ASR (Automatic Speech Recognition) service to recognize the event. The platform then sends the recognition result to the NLU (Natural Language Understanding) service to determine the intent. Based on the intent, the platform sends a directive (e.g., "Play" directive) to the Skills service (e.g., Music Service). The Skills service then sends a Text/SSML response to the TTS (Text-to-Speech) service, which generates a streaming URL. The platform then plays back the response through a speaker. Below the diagram, the text "What do events and directives look like?" is followed by an example of an event JSON payload. The payload is a JSON object with a header, a payload, and a context. The context is a JSON object with a header, a namespace, a name, a payload, and a token. The token is a string representing the playback state. The payload is a JSON object with a token, a player activity, and an offset in milliseconds.</p> <pre> SpeechRecognizer.Recognize 2017-10-10 18:19:25 { "event": { "header": { "dialogRequestId": "dialogRequestId-2fa1867c-bfe2-4113-9be0-325a1c466156", "namespace": "SpeechRecognizer", "name": "Recognize", "messageId": "e4717a2e-16bb-4153-ae31-3068107542be" }, "payload": { "profile": "CLOSE_TALK", "format": "AUDIO_L16_RATE_16000_CHANNELS_1" } }, "context": { "header": { "namespace": "AudioPlayer", "name": "Playbackstate" }, "payload": { "token": "", "playerActivity": "TALK", "offsetInMilliseconds": "0" } } } </pre> <p>Event</p> <p>AWS re:Invent © 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.</p>

AWS re: Invent, Integrate Alexa into Your Product, Using the Alexa Voice Service Device SDK.

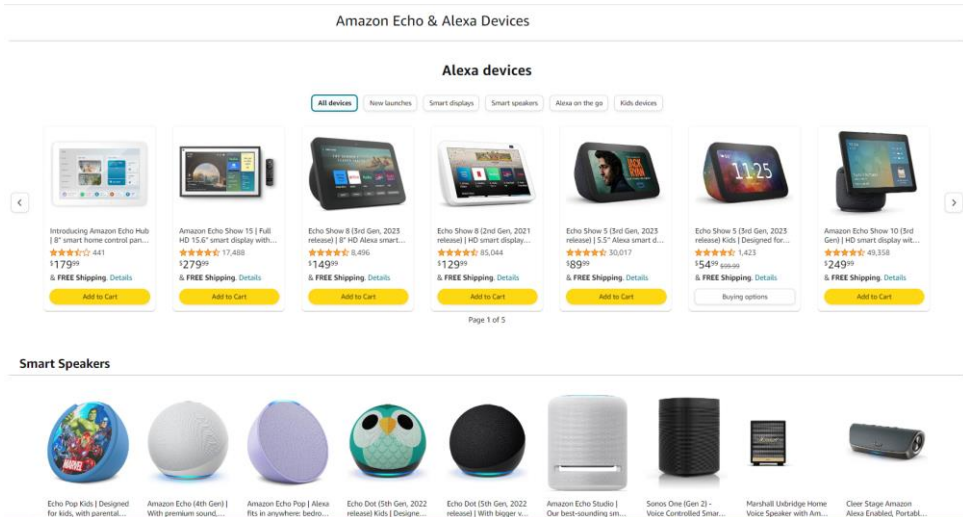
Claim	Exemplary Disclosures
	  <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y</p>
<p>[1.h] determining, by the computer system, based on the first context, an interpretation of the first natural language utterance; and</p>	<p>Alexa is programmed to perform and have performed methods of determining, by the computer system, based on the first context, an interpretation of the first natural language utterance.</p> <p>Alexa determines that based on the context the first natural language means to choose the video to play from a list of results.</p>

Claim	Exemplary Disclosures
	<p data-bbox="720 248 1604 272">Digital Services and Device Support › Alexa Features Help › Video with Alexa ›</p> <h2 data-bbox="720 285 1583 334">Play Videos with Alexa Using Your Voice</h2> <p data-bbox="720 376 1623 443">Use your voice to play movies, videos, or TV shows from Prime Video and supported streaming services on Alexa devices and Fire TV devices.</p> <p data-bbox="699 483 861 508">Video Search*</p> <p data-bbox="699 534 1703 592">After searching for video content, say "Select [video]" or "Play [number]" to choose a video from the results.</p> <p data-bbox="699 630 1136 654">Actions on Detail Page* (Fire TV Only)</p> <ul data-bbox="720 690 1041 784" style="list-style-type: none"> • "Play this." • "Add this to my Watchlist." • "Play the trailer." <p data-bbox="699 812 1713 870">Note: * Not all features are supported from a paired Echo Show or for a paired Fire TV Cube (2nd Generation).</p> <p data-bbox="686 907 1925 938">See https://www.amazon.com/gp/help/customer/display.html?nodeId=G3YFGTLKXC2YWLQ</p>
[1.i] generating, by the computer system, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.	<p data-bbox="686 976 1904 1076">Alexa is programmed to perform and have performed methods of generating, by the computer system, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.</p> <p data-bbox="686 1117 1518 1148">For example, Alexa will generate a verbal response immediately.</p>

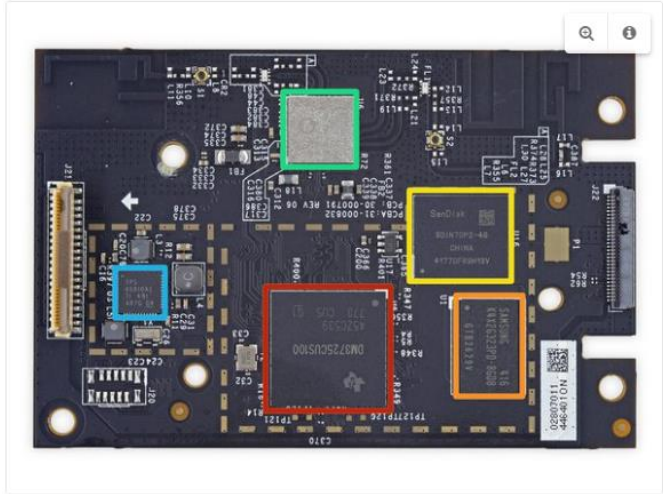
Claim	Exemplary Disclosures
	 <p data-bbox="684 753 1381 789">See https://www.youtube.com/watch?v=IjOcw1utv4A</p>  <p data-bbox="684 1378 1407 1414">See https://www.youtube.com/watch?v=N6NWp8_IzZ8</p>

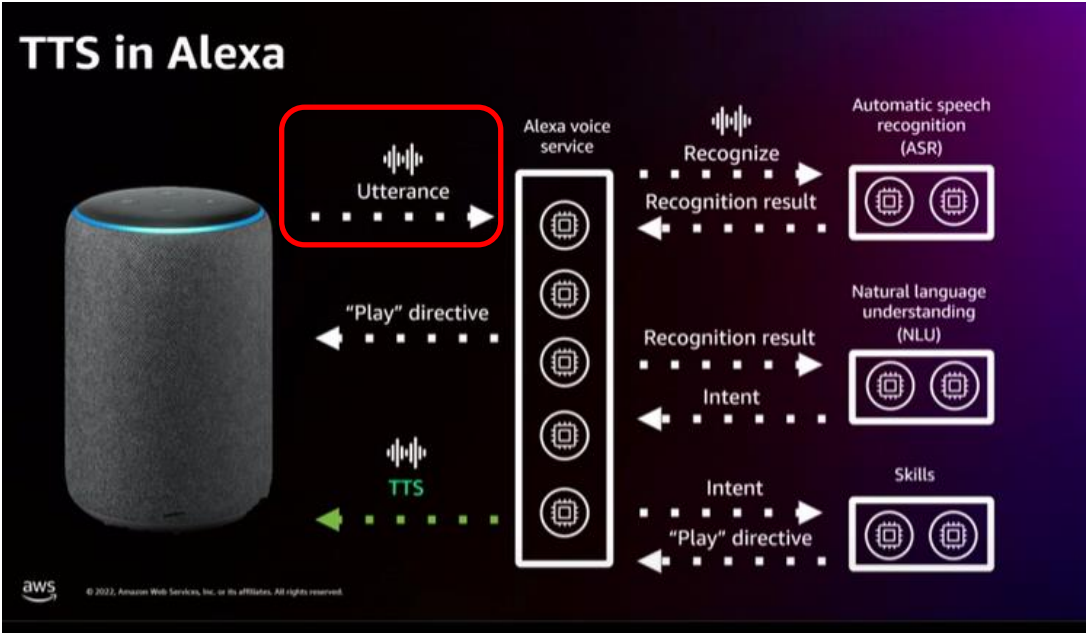
EXHIBIT G

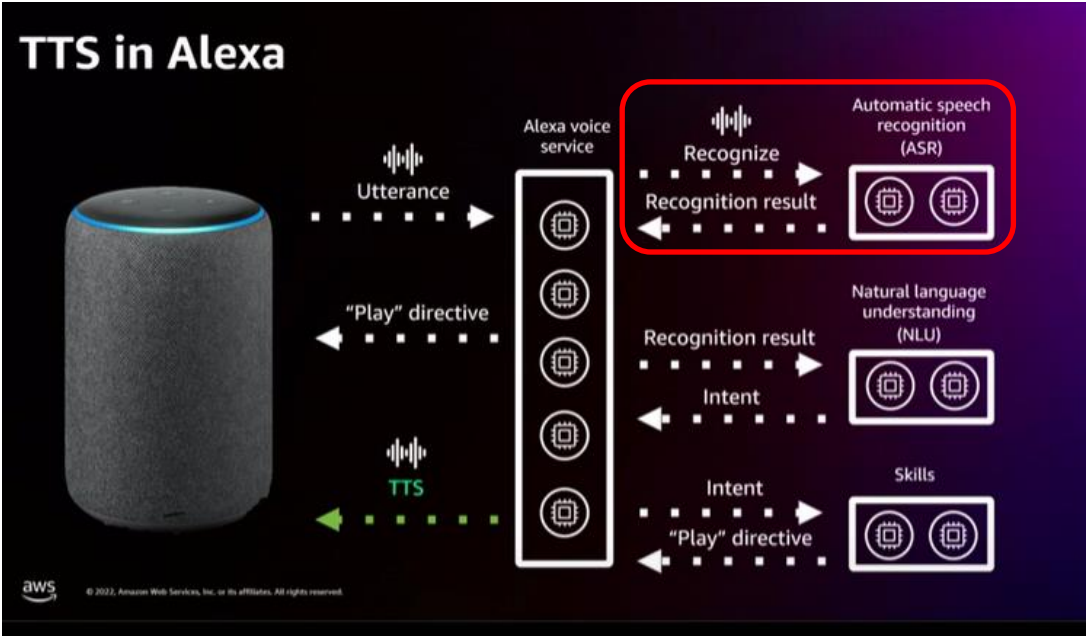
Exhibit G: Exemplary Infringement Claim Chart for U.S. Patent 10,755,699 Claim 1

Claim	Exemplary Disclosures
<p>[1.pre] A computer-implemented method of generating natural language system responses adapted based on a user's manner of speaking, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:</p>	<p>Amazon's Alexa is programmed to perform and have performed methods for generating natural language system responses adapted based on a user's manner of speaking, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method.</p> <p>Amazon's Alexa includes for example, Echo, Echo Dot, Echo Show, FireTV, Alexa apps, and Alexa cloud).</p>  <p>See https://www.amazon.com/b?&node=9818047011&ref=MARS_NAV_desktop_plswitcher</p> <p>Alexa users can use the voice purchasing function.</p>

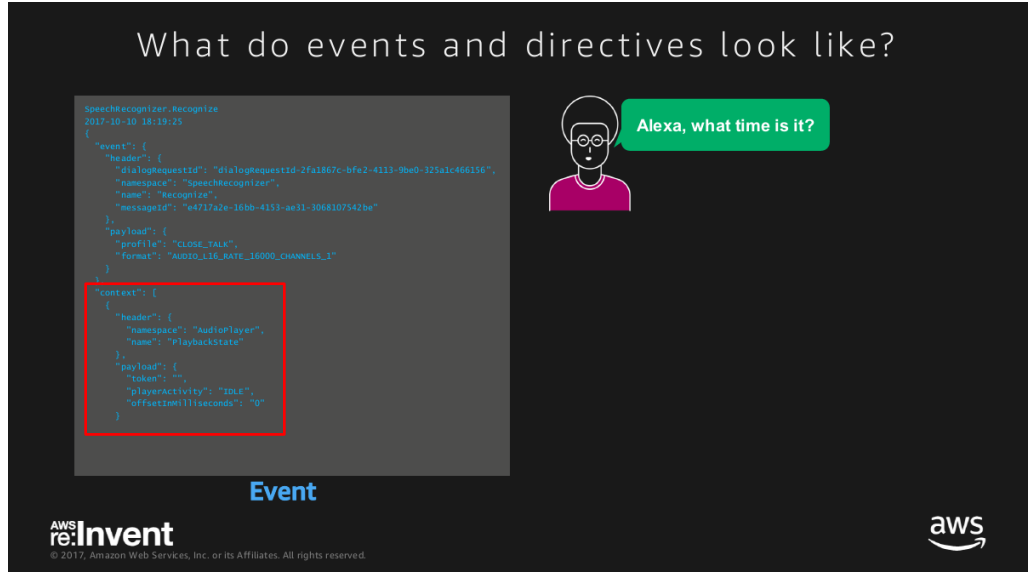
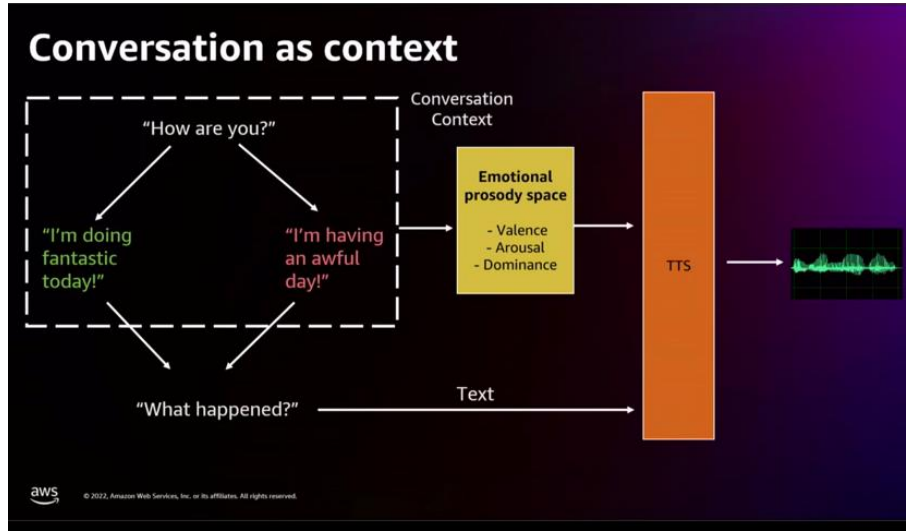
Claim	Exemplary Disclosures
	<p data-bbox="709 250 1367 272">Digital Services and Device Support › Alexa Features Help › Shopping with Alexa ›</p> <h2 data-bbox="709 277 1297 313">Turn Alexa Voice Purchasing On or Off</h2> <p data-bbox="709 342 1327 365">To turn Voice Purchasing on or off for your account, use the Alexa app.</p> <hr data-bbox="709 381 1436 384"/> <ol data-bbox="724 402 1026 521" style="list-style-type: none"> 1. Open the Alexa app . 2. Open More  and select Settings. 3. Select Account Settings. 4. Select Voice Purchasing. 5. Turn Voice Purchasing on or off. <p data-bbox="709 542 1421 586">Note: You can limit purchasing so that only recognized household members can place orders with Alexa.</p> <p data-bbox="709 607 861 630">Related Help Topics</p> <p data-bbox="709 651 1098 673">Manage an Alexa Voice ID for Purchases with Alexa</p> <div data-bbox="709 690 1192 805"> <p>Was this information helpful?</p> <div> <input data-bbox="737 753 779 786" type="button" value="Yes"/> <input data-bbox="793 753 835 786" type="button" value="No"/> </div> </div> <p data-bbox="688 870 1871 906">https://www.amazon.com/gp/help/customer/display.html?nodeId=GPUCQ6PMPMENG8FA</p> <p data-bbox="688 943 1902 1044">Additionally, the voice purchasing function is implemented on Amazon's Alexa that has one or more physical processors. See https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/</p> <p data-bbox="688 1081 1885 1149">Additionally, Echo devices have one or more processors. For example, the Texas Instruments DM3725CUS100 Digital Media Processor as depicted below with red annotations:</p>

Claim	Exemplary Disclosures
	 <p data-bbox="688 722 1554 792">Sam Goldheart, Amazon Echo Teardown, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953</p>
[1.a] receiving, by the computer system, a user input comprising a natural language utterance;	Alexa is programmed to perform and have performed methods of receiving, by the computer system, a user input comprising a natural language utterance.

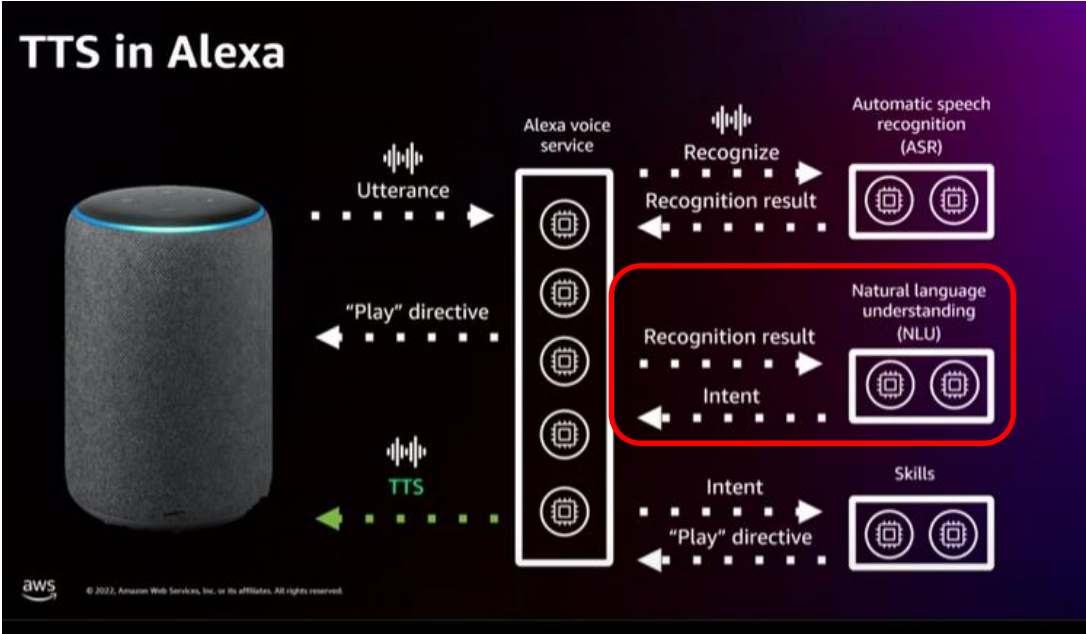
Claim	Exemplary Disclosures
	 <p>The diagram, titled "TTS in Alexa", illustrates the process flow for a voice assistant. It features an Alexa device on the left. An "Utterance" (represented by a sound wave icon and a red box) is sent to the "Alexa voice service" (a vertical column of four processor icons). From the Alexa voice service, the process flows to three main components: "Automatic speech recognition (ASR)", "Natural language understanding (NLU)", and "Skills". Each component has a "Recognition result" or "Intent" step that feeds back into the Alexa voice service. The "Skills" component also has a "Play directive" step that feeds back into the Alexa voice service. Finally, a "TTS" (Text-to-Speech) step (represented by a green sound wave icon) is shown at the bottom, leading back to the Alexa device. The AWS logo and copyright notice are at the bottom left of the diagram.</p> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
[1.b] recognizing, by the computer system, one or more words or phrases from the natural language utterance;	Alexa is programmed to perform and have performed methods of recognizing, by the computer system, one or more words or phrases from the natural language utterance.

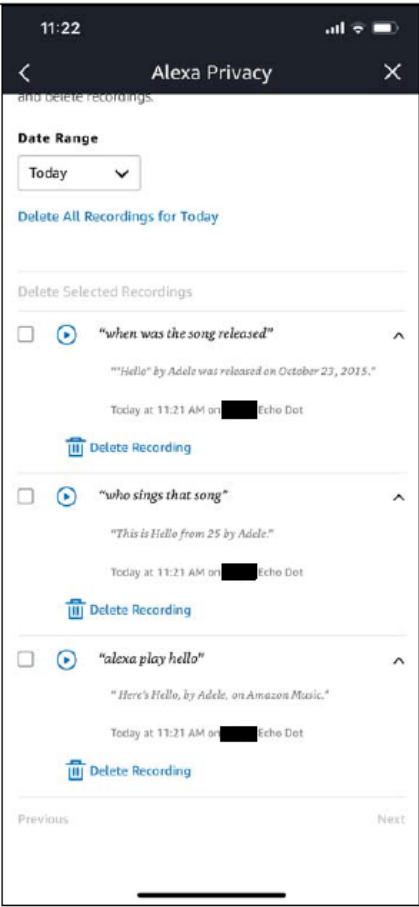
Claim	Exemplary Disclosures
	 <p>The diagram, titled "TTS in Alexa", illustrates the process of handling a user's voice command. It shows an Alexa device on the left. An "Utterance" (represented by a soundwave icon) is sent from the device to the "Alexa voice service" (a vertical stack of four processor icons). From the voice service, the data flows to the "Automatic speech recognition (ASR)" module. The ASR module sends a "Recognition result" to the "Natural language understanding (NLU)" module. The NLU module sends an "Intent" to the "Skills" module. The Skills module sends a "Play" directive back to the Alexa voice service. Finally, the voice service uses "TTS" (Text-to-Speech, indicated by a green soundwave icon) to send a response back to the Alexa device. The ASR and NLU modules are each represented by a box containing two processor icons. The Skills module is represented by a box containing two processor icons and a play button icon. A red box highlights the ASR module and its connection to the NLU module.</p> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
<p>[1.c] identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;</p>	<p>Alexa is programmed to perform and have performed methods of identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance.</p> <p>“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p>“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p>

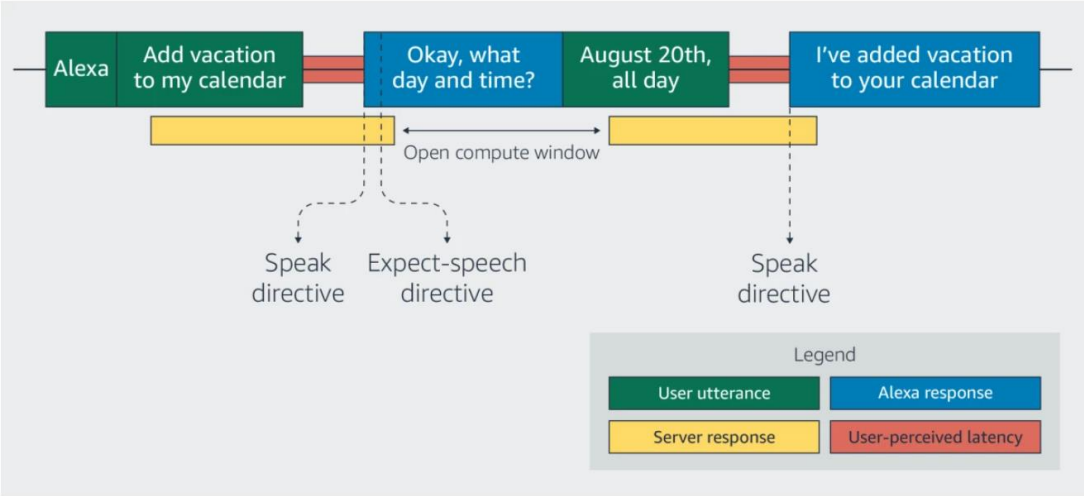
Claim	Exemplary Disclosures														
	<p>https://www.amazon.science/latest-news/the-engineering-behind-alexa-contextual-speech-recognition</p> <h3 data-bbox="821 386 1142 427">Contextual carryover</h3> <p data-bbox="821 464 1818 602">You don't need to provide annotated dialogs that demonstrate contextual carryover. The dialog simulator supports this feature by introducing turns with pronouns (for example, "Please purchase it") when generating dialog variants from annotated dialogs. The carryover occurs at a later runtime stage, argument filling, when Alexa Conversations considers all slots that the user and Alexa mention across the entire dialog for filling API arguments.</p> <p data-bbox="821 634 903 659">Example</p> <p data-bbox="821 688 1719 712">You provide a dialog to purchase a movie. The dialog simulator introduces a turn with a pronoun.</p> <table data-bbox="821 743 1818 987"> <thead> <tr> <th data-bbox="821 743 1297 792">Dialog you provide</th><th data-bbox="1297 743 1818 792">Example variant from the dialog simulator</th></tr> </thead> <tbody> <tr> <td data-bbox="821 805 1297 829">User: I'd like to watch a movie from Guy Ritchie.</td><td data-bbox="1297 805 1818 829">User: I'd like to watch a movie from Guy Ritchie.</td></tr> <tr> <td data-bbox="821 834 1297 859">(Invoke API RecommendMovie.)</td><td data-bbox="1297 834 1818 859">(Invoke API RecommendMovie.)</td></tr> <tr> <td data-bbox="821 863 1297 888">Alexa: How about Snatch?</td><td data-bbox="1297 863 1818 888">Alexa: How about Snatch?</td></tr> <tr> <td data-bbox="821 893 1297 917">User: Please help me purchase Snatch.</td><td data-bbox="1297 893 1818 917">User: Please purchase it.</td></tr> <tr> <td data-bbox="821 922 1297 946">(Invoke API ReserveMovie.)</td><td data-bbox="1297 922 1818 946">(Invoke API ReserveMovie.)</td></tr> <tr> <td data-bbox="821 951 1297 976">Alexa: The movie Snatch is reserved for you.</td><td data-bbox="1297 951 1818 976">Alexa: The movie Snatch is reserved for you.</td></tr> </tbody> </table> <h3 data-bbox="821 1029 1079 1070">Context memory</h3> <p data-bbox="821 1107 1818 1219"><i>Context memory</i> is short-term memory that tracks user utterances and models results such as entities (that is, a slot that the user mentioned or a return value from a previous API), predicted actions, and Alexa responses. Context memory maintains dialog context for each session and is constantly updated throughout the session.</p> <p data-bbox="688 1243 1881 1308">See https://developer.amazon.com/en-US/docs/alexa/conversations/how-alexa-conversations-works.html#contextual-carryover</p>	Dialog you provide	Example variant from the dialog simulator	User: I'd like to watch a movie from Guy Ritchie.	User: I'd like to watch a movie from Guy Ritchie.	(Invoke API RecommendMovie.)	(Invoke API RecommendMovie.)	Alexa: How about Snatch?	Alexa: How about Snatch?	User: Please help me purchase Snatch.	User: Please purchase it.	(Invoke API ReserveMovie.)	(Invoke API ReserveMovie.)	Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.
Dialog you provide	Example variant from the dialog simulator														
User: I'd like to watch a movie from Guy Ritchie.	User: I'd like to watch a movie from Guy Ritchie.														
(Invoke API RecommendMovie.)	(Invoke API RecommendMovie.)														
Alexa: How about Snatch?	Alexa: How about Snatch?														
User: Please help me purchase Snatch.	User: Please purchase it.														
(Invoke API ReserveMovie.)	(Invoke API ReserveMovie.)														
Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.														

Claim	Exemplary Disclosures
	<div data-bbox="682 212 1703 781"> <h3>What do events and directives look like?</h3> <pre> SpeechRecognizer.Recognize 0017-10-10 18:19:25 { "event": { "header": { "dialogRequestId": "2fa1867c-bfe2-4113-9bd0-325a1c466156", "namespace": "SpeechRecognizer", "name": "recognize", "messageId": "e4717a2e-16bb-4153-ae31-3068107542be" }, "payload": { "profile": "CLOSE_TALK", "format": "AUDIO_L16_RATE_16000_CHANNELS_1" } }, "context": { "header": { "namespace": "AudioPlayer", "name": "PlaybackState" }, "payload": { "token": "", "playerActivity": "IDLE", "offsetInMilliseconds": "0" } } } </pre> <p style="text-align: center;">Event</p>  </div> <p>AWS re: Invent, Integrate Alexa into Your Product, Using the Alexa Voice Service Device SDK.</p> <div data-bbox="682 919 1581 1446"> <h3>Conversation as context</h3>  </div>

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y</p>
<p>[1.d] determining, by the computer system, an interpretation of the natural language utterance based on the identified context;</p>	<p>The Accused Products are programmed to perform and have performed methods of determining, by the computer system, an interpretation of the natural language utterance based on the identified context.</p> <p>“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p>“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p> <p>https://www.amazon.science/latest-news/the-engineering-behind-alexas-contextual-speech-recognition</p>

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y at 10:07.</p>
<p>[1.e] accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;</p>	<p>Alexa is programmed to perform and have performed methods of accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system.</p> <p>An example is accumulating short-term shared knowledge about the current conversation shown below in a screenshot of Alexa app’s “Review Voice History” page below.</p>

Claim	Exemplary Disclosures
	 <p>The screenshot displays the 'Alexa Privacy' interface on a mobile device. At the top, the status bar shows the time as 11:22 and signal/battery icons. The app header includes a back arrow, the title 'Alexa Privacy', and a close 'X' button. Below the header, the text 'and delete recordings.' is partially visible. A 'Date Range' section features a dropdown menu set to 'Today'. A blue link 'Delete All Recordings for Today' is present. A horizontal separator is followed by the heading 'Delete Selected Recordings'. Three recordings are listed, each with a selection checkbox, a play button icon, a transcript snippet, a timestamp, the device name 'Echo Dot', and a 'Delete Recording' button with a trash icon. The first recording is titled 'when was the song released?' with the transcript 'Hello by Adele was released on October 23, 2015.' The second is 'who sings that song?' with 'This is Hello from 25 by Adele.' The third is 'alexa play hello' with 'Here's Hello, by Adele, on Amazon Music.' At the bottom, 'Previous' and 'Next' navigation options are shown.</p>

Claim	Exemplary Disclosures
	 <p>Because the instruction to re-open the microphone (<i>expect-speech directive</i>) follows immediately upon the instruction to begin executing Alexa's reply (<i>speak directive</i>), the reply itself buys the context model enough time to produce a context vector.</p> <p>See https://www.amazon.science/latest-news/the-engineering-behind-alexas-contextual-speech-recognition</p> <p>The accumulation occurs during a predetermined time period and are related to a single conversation between Alexa and the user.</p> <p>“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p>“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p> <p>https://www.amazon.science/latest-news/the-engineering-behind-alexas-contextual-speech-recognition</p>

Claim	Exemplary Disclosures										
	<p data-bbox="821 282 1142 321">Contextual carryover</p> <p data-bbox="821 362 1818 500">You don't need to provide annotated dialogs that demonstrate contextual carryover. The dialog simulator supports this feature by introducing turns with pronouns (for example, "Please purchase it") when generating dialog variants from annotated dialogs. The carryover occurs at a later runtime stage, argument filling, when Alexa Conversations considers all slots that the user and Alexa mention across the entire dialog for filling API arguments.</p> <p data-bbox="821 529 903 553">Example</p> <p data-bbox="821 584 1719 609">You provide a dialog to purchase a movie. The dialog simulator introduces a turn with a pronoun.</p> <table data-bbox="821 638 1818 883"> <thead> <tr> <th data-bbox="829 651 1297 686">Dialog you provide</th><th data-bbox="1325 651 1810 686">Example variant from the dialog simulator</th></tr> </thead> <tbody> <tr> <td data-bbox="829 703 1297 756">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td><td data-bbox="1325 703 1810 756">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td></tr> <tr> <td data-bbox="829 760 1297 784">Alexa: How about Snatch?</td><td data-bbox="1325 760 1810 784">Alexa: How about Snatch?</td></tr> <tr> <td data-bbox="829 787 1297 841">User: Please help me purchase Snatch. (Invoke API ReserveMovie.)</td><td data-bbox="1325 787 1810 841">User: Please purchase it. (Invoke API ReserveMovie.)</td></tr> <tr> <td data-bbox="829 844 1297 868">Alexa: The movie Snatch is reserved for you.</td><td data-bbox="1325 844 1810 868">Alexa: The movie Snatch is reserved for you.</td></tr> </tbody> </table> <p data-bbox="821 924 1079 963">Context memory</p> <p data-bbox="821 1003 1810 1109"><i>Context memory</i> is short-term memory that tracks user utterances and models results such as entities (that is, a slot that the user mentioned or a return value from a previous API), predicted actions, and Alexa responses. Context memory maintains dialog context for each session and is constantly updated throughout the session.</p> <p data-bbox="688 1138 1881 1203">See https://developer.amazon.com/en-US/docs/alexa/conversations/how-alexa-conversations-works.html#contextual-carryover</p>	Dialog you provide	Example variant from the dialog simulator	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	Alexa: How about Snatch?	Alexa: How about Snatch?	User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)	Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.
Dialog you provide	Example variant from the dialog simulator										
User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)										
Alexa: How about Snatch?	Alexa: How about Snatch?										
User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)										
Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.										

Claim	Exemplary Disclosures
	<div data-bbox="688 212 1703 781"> <h3>What do events and directives look like?</h3> <div style="display: flex; align-items: center;"> <pre style="background-color: #333; color: #eee; padding: 10px; margin-right: 20px;"> SpeechRecognizer.Recognize 0017-10-10 18:19:25 { "event": { "header": { "dialogRequestId": "2fa1867c-bfe2-4113-9be0-325a1c466156", "namespace": "SpeechRecognizer", "name": "recognize", "messageId": "e4717a2e-16bb-4153-ae31-3068107542be" }, "payload": { "profile": "CLOSE_TALK", "format": "AUDIO_L16_RATE_16000_CHANNELS_1" } }, "context": { "header": { "namespace": "AudioPlayer", "name": "PlaybackState" }, "payload": { "token": "", "playerActivity": "IDLE", "offsetInMilliseconds": "0" } } } </pre> <div style="text-align: center;">  <p style="background-color: #00a651; color: white; padding: 5px; border-radius: 10px; display: inline-block;">Alexa, what time is it?</p> </div> </div> <p style="text-align: center; color: #00a651; font-weight: bold; margin-top: 10px;">Event</p> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 10px;">   </div> <p style="font-size: 0.8em; margin-top: 5px;">© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.</p> </div> <p style="margin-top: 20px;">AWS re: Invent, Integrate Alexa into Your Product, Using the Alexa Voice Service Device SDK.</p>

Claim	Exemplary Disclosures
	<div data-bbox="688 212 1703 808"> <h3>Conversation as context</h3> <p>© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.</p> </div> <div data-bbox="688 813 1703 1365"> <h3>Adapting speech to the conversation</h3> <p>Alexa, help me find some music.</p> <p>Happy to help. Let's find something you'll like on Amazon Music. Any particular type of music you wanted?</p> <p>Let's have some rock music.</p> <p>Got it. How about something by The Eagles?</p> <p>I'm not a big fan of The Eagles.</p> <p>Hmm, okay. How about we try some songs by Queen?</p> <p>No, give me something more modern!</p> <p>Oh dear, I'm not getting you at all today. What would you think about something by Foo Fighters?</p> <p>That's more like it.</p> <p>Great. In that case, do you think you'd like Vampire Weekend?</p> <p>Now you're getting it . . .</p> <p>Fantastic. Pulling a playlist together for you right now.</p> <p>© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.</p> </div> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y</p>

Claim	Exemplary Disclosures
	<p>Alexa also accumulates short-term knowledge in determining whether users' communicated in normal voice or whisper.</p> <p>"LSTMs are widely used in speech recognition and natural-language understanding because they process inputs in sequential order, and their judgments about any given input are conditioned by what they've already seen."</p> <p>In a process called "end-pointing", however, Alexa recognizes the end of an utterance by the short period of silence that follows end of speech, and that silence is part of the input to the whisper detector."</p> <p>https://www.amazon.science/blog/whisper-to-alexa-and-shell-whisper-back?tag=undefined</p>
<p>[1.f] accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;</p>	<p>Alexa is programmed to perform and have performed methods of accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period.</p> <p>An example of accumulating long-term knowledge is based on user profiles, as shown below.</p> <p>"You can create a voice ID so Alexa can call you by name and do more to personalize your experience. When you create a voice ID, Alexa uses recordings of your voice to create an acoustic model of your voice characteristics and to update that model over time to improve Alexa's ability to recognize you."</p> <p>https://www.amazon.com/gp/help/customer/display.html?nodeId=201809740</p> <p>"Voice ID analyzes speech attributes like rhythm, pitch, and tone to create a digital voiceprint,"</p> <p>https://aws.amazon.com/connect/voice-id/#:~:text=When%20a%20customer%20agrees%20to,customer%20on%20their%20next%20call.</p>

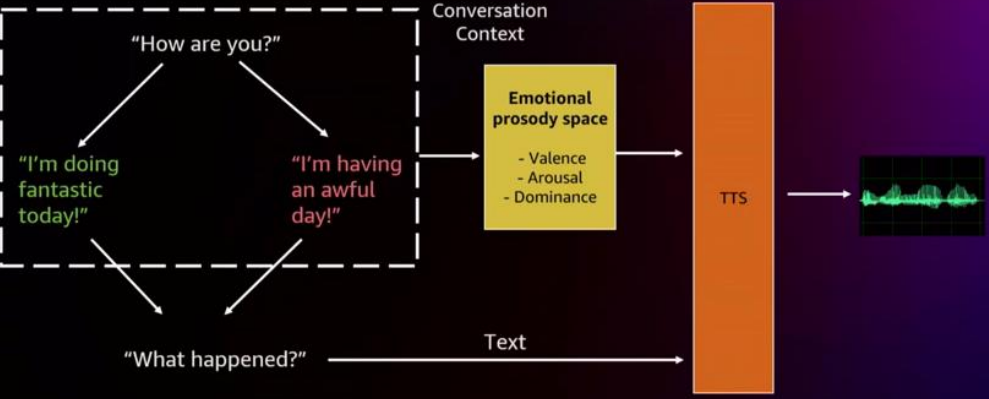
Claim	Exemplary Disclosures
	<p data-bbox="716 224 1386 245">Digital Services and Device Support › Alexa Settings › Alexa Profile Settings ›</p> <h2 data-bbox="716 253 1100 289">What Is Alexa Voice ID?</h2> <p data-bbox="716 321 1386 370">Alexa voice ID helps Alexa recognize you when you speak and provide a personalized experience.</p> <hr data-bbox="716 394 1499 397"/> <p data-bbox="716 414 1486 483">You can create an Alexa voice ID for a personalized experience when Alexa recognizes your voice. With Alexa voice ID, Alexa can call you by name and provide enhanced personalization. To set up Alexa voice ID, go to Create an Alexa Voice ID.</p> <p data-bbox="716 511 1257 532">With an Alexa voice ID, you can get a personalized experience for:</p> <ul data-bbox="730 557 1493 914" style="list-style-type: none"> • Alexa Calling and Messaging: Send and receive your own personal messages. • Calendar: Hear your personal upcoming schedule and add events to your calendar. • Flash Briefing: Skip stories and news items that you've already heard. • Guest Connect: Connect your account to a friend's eligible Echo device. Once connected, Alexa uses voice recognition to provide you with access to your music and news when you're away from home. • Music: If you have the Amazon Music Unlimited Family Plan, Alexa customizes music playback based on your voice. • Music and Podcast Services: Set default music and podcast services for each Alexa profile on your Alexa account. • Reminders: Create personal reminders and get them delivered just to your phone. • Shopping: Limit purchasing so that only recognized household members can place orders with Alexa. • Skills: Receive personalized responses for supported skills. <p data-bbox="716 935 1415 979">Voice ID uses voice recognition to recognize you and display personalized content on compatible Echo Show devices.</p> <p data-bbox="716 1003 882 1024">Related Help Topics</p> <p data-bbox="716 1049 1037 1092"> Create an Alexa Voice ID Create an Alexa Voice ID for Your Child </p> <p data-bbox="688 1130 1877 1195"> <i>See</i> https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXYKY2AB2QWZT2X </p> <p data-bbox="688 1232 1913 1369"> “To be as useful as possible to customers, Alexa should be able to make educated guesses about the meanings of ambiguous utterances. If, for instance, a customer says, “Alexa, play the song ‘Hello’”, Alexa should be able to infer from the customer’s listening history whether the song requested is the one by Adele or the one by Lionel Richie.” </p> <p data-bbox="688 1406 1927 1471"> “We use machine learning to analyze playback duration data to infer song preference, and we use collaborative-filtering techniques to estimate how a particular customer might rate a song that he </p>

Claim	Exemplary Disclosures
	<p>or she has never requested. Although we tested our approach on music-streaming records, it generalizes easily to any other streaming-data service, such as video or audiobooks.”</p> <p>“Customers’ ratings for all the songs in a music service’s catalogue could be represented as an enormous grid, or matrix. Each row represents a customer, each column represents a song, and the value entered at the intersection of any row and any column indicates the customer’s “affinity” for a particular song.”</p> <p>See https://www.amazon.science/blog/how-alexa-can-use-song-playback-duration-to-learn-customers-preferences</p> <p>“You can enroll your Amazon Profile in Alexa as an Alexa Profile. Alexa Profiles gives each person in your home their own individual Alexa experience. You can create an Alexa Profile for you and your family members when setting up your device. Alexa Profiles can be created from your device after setup. Select an existing Amazon Profile on your Amazon Account or create a new Amazon Profile. Existing Amazon Profiles include profiles enrolled in other programs, like Prime Video.”</p> <p>https://www.amazon.com/gp/help/customer/display.html?nodeId=GLE4EVA2VLDRQDLC</p>
[1.g] identifying, by the computer system, a manner in which the natural language utterance was spoken based the short-term knowledge and the long-term knowledge; and	<p>Alexa is programmed to perform and have performed methods of identifying, by the computer system, a manner in which the natural language utterance was spoken based the short-term knowledge and the long-term knowledge.</p>

Claim	Exemplary Disclosures
	<p data-bbox="716 224 1388 245">Digital Services and Device Support › Alexa Settings › Alexa Profile Settings ›</p> <h2 data-bbox="716 253 1100 289">What Is Alexa Voice ID?</h2> <p data-bbox="716 321 1388 370">Alexa voice ID helps Alexa recognize you when you speak and provide a personalized experience.</p> <hr data-bbox="716 394 1499 397"/> <p data-bbox="716 414 1486 483">You can create an Alexa voice ID for a personalized experience when Alexa recognizes your voice. With Alexa voice ID, Alexa can call you by name and provide enhanced personalization. To set up Alexa voice ID, go to Create an Alexa Voice ID.</p> <p data-bbox="716 511 1260 532">With an Alexa voice ID, you can get a personalized experience for:</p> <ul data-bbox="730 557 1493 914" style="list-style-type: none"> • Alexa Calling and Messaging: Send and receive your own personal messages. • Calendar: Hear your personal upcoming schedule and add events to your calendar. • Flash Briefing: Skip stories and news items that you've already heard. • Guest Connect: Connect your account to a friend's eligible Echo device. Once connected, Alexa uses voice recognition to provide you with access to your music and news when you're away from home. • Music: If you have the Amazon Music Unlimited Family Plan, Alexa customizes music playback based on your voice. • Music and Podcast Services: Set default music and podcast services for each Alexa profile on your Alexa account. • Reminders: Create personal reminders and get them delivered just to your phone. • Shopping: Limit purchasing so that only recognized household members can place orders with Alexa. • Skills: Receive personalized responses for supported skills. <p data-bbox="716 935 1417 979">Voice ID uses voice recognition to recognize you and display personalized content on compatible Echo Show devices.</p> <p data-bbox="716 1003 882 1024">Related Help Topics</p> <p data-bbox="716 1049 1037 1092"> Create an Alexa Voice ID Create an Alexa Voice ID for Your Child </p> <p data-bbox="688 1130 1877 1195"> See https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXYKY2AB2QWZT2X </p> <p data-bbox="688 1252 1940 1279">Additionally, Alexa can identify whether the utterance was spoken in a whisper or a normal voice.</p> <p data-bbox="688 1320 1946 1455">“Whispered speech is predominantly unvoiced, meaning that it doesn’t involve the vibration of the vocal cords, and it has less energy in lower frequency bands than ordinary speech. Previously, researchers have sought to exploit these facts by training their classifiers, not on raw speech signals, but on “features” extracted from the signals, which are designed to capture information</p>

Claim	Exemplary Disclosures
	<p>that could help discriminate whispers from normal speech.” See https://www.amazon.science/blog/whisper-to-alexa-and-shell-whisper-back?tag=undefined</p> <p>“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p>“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p> <p>https://www.amazon.science/latest-news/the-engineering-behind-alexas-contextual-speech-recognition</p>

Claim	Exemplary Disclosures										
	<p>Contextual carryover</p> <p>You don't need to provide annotated dialogs that demonstrate contextual carryover. The dialog simulator supports this feature by introducing turns with pronouns (for example, "Please purchase it") when generating dialog variants from annotated dialogs. The carryover occurs at a later runtime stage, argument filling, when Alexa Conversations considers all slots that the user and Alexa mention across the entire dialog for filling API arguments.</p> <p>Example</p> <p>You provide a dialog to purchase a movie. The dialog simulator introduces a turn with a pronoun.</p> <table border="1"> <thead> <tr> <th data-bbox="821 602 1297 651">Dialog you provide</th><th data-bbox="1297 602 1814 651">Example variant from the dialog simulator</th></tr> </thead> <tbody> <tr> <td data-bbox="821 651 1297 691">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td><td data-bbox="1297 651 1814 691">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td></tr> <tr> <td data-bbox="821 691 1297 716">Alexa: How about Snatch?</td><td data-bbox="1297 691 1814 716">Alexa: How about Snatch?</td></tr> <tr> <td data-bbox="821 716 1297 756">User: Please help me purchase Snatch. (Invoke API ReserveMovie.)</td><td data-bbox="1297 716 1814 756">User: Please purchase it. (Invoke API ReserveMovie.)</td></tr> <tr> <td data-bbox="821 756 1297 797">Alexa: The movie Snatch is reserved for you.</td><td data-bbox="1297 756 1814 797">Alexa: The movie Snatch is reserved for you.</td></tr> </tbody> </table> <p>Context memory</p> <p><i>Context memory</i> is short-term memory that tracks user utterances and models results such as entities (that is, a slot that the user mentioned or a return value from a previous API), predicted actions, and Alexa responses. Context memory maintains dialog context for each session and is constantly updated throughout the session.</p> <p>See https://developer.amazon.com/en-US/docs/alexa/conversations/how-alexa-conversations-works.html#contextual-carryover</p>	Dialog you provide	Example variant from the dialog simulator	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	Alexa: How about Snatch?	Alexa: How about Snatch?	User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)	Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.
Dialog you provide	Example variant from the dialog simulator										
User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)										
Alexa: How about Snatch?	Alexa: How about Snatch?										
User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)										
Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.										

Claim	Exemplary Disclosures
	<div data-bbox="688 212 1703 781"> <h3>What do events and directives look like?</h3> <div data-bbox="751 305 1186 683"> <pre>SpeechRecognizer.Recognize 0017-10-10 18:19:25 { "event": { "header": { "dialogRequestId": "2fa1867c-bfa2-4113-9bd0-325a1c466156", "namespace": "SpeechRecognizer", "name": "recognize", "messageId": "e4717a2e-16bb-4153-ae31-3068107542be" }, "payload": { "profile": "CLOSE_TALK", "format": "AUDIO_16K_RATE_16000_CHANNELS_1" } }, "context": { "header": { "namespace": "AudioPlayer", "name": "PlaybackState" }, "payload": { "token": "", "playerActivity": "IDLE", "offsetInMilliseconds": "0" } } }</pre> <p>Event</p> </div> <div data-bbox="1220 305 1507 423">  <p>Alexa, what time is it?</p> </div> <div data-bbox="730 727 856 776"> <p>AWS re:Invent © 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.</p> </div> <div data-bbox="1598 727 1661 776"> <p>aws</p> </div> </div> <p>AWS re: Invent, Integrate Alexa into Your Product, Using the Alexa Voice Service Device SDK.</p> <div data-bbox="688 850 1692 1442"> <h3>Conversation as context</h3>  <p>© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.</p> </div>

Claim	Exemplary Disclosures
	 <p>The screenshot shows a conversation with Alexa on a dark background. The user asks for music, then specifies rock music, then expresses dislike for The Eagles, and finally asks for something more modern. Alexa responds with suggestions like Queen and Foo Fighters, and finally creates a playlist. The AWS logo and copyright notice are at the bottom of the screenshot.</p> <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y</p>
<p>[1.h] generating, by the computer system, a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.</p>	<p>Alexa is programmed to perform and have performed methods of generating, by the computer system, a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.</p> <p>For example, the response may be impacted by the identity of the user, which is ascertained by the manner in which the utterance was spoken. The response may also be impacted by whether the end-user's utterance was at a whisper or a normal voice.</p> <p>An example is generating a response based on the identified manner. Alexa: "Resume my music." In response, Alexa would resume a song that had been paused with respect to the particular VoiceID for the specified device. Or asking Alexa for a selection of available audio books, which is based on the user's VoiceID.</p> <p>Another example would be any response that Alexa make using a whisper voice, since a whisper is in response to the user whispering.</p>

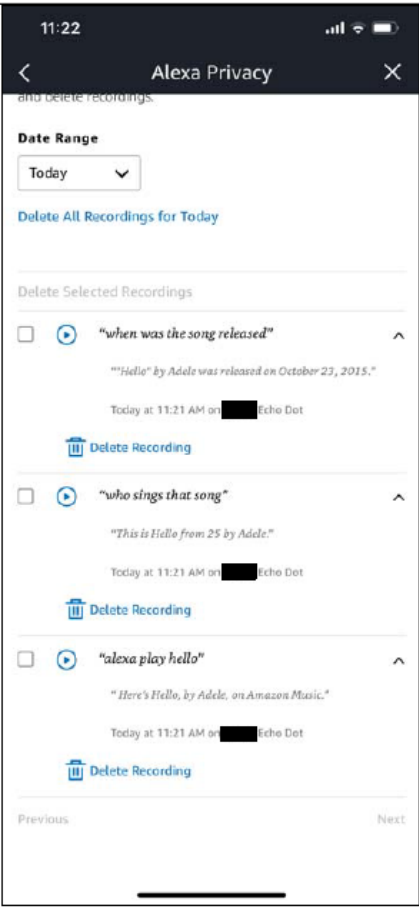
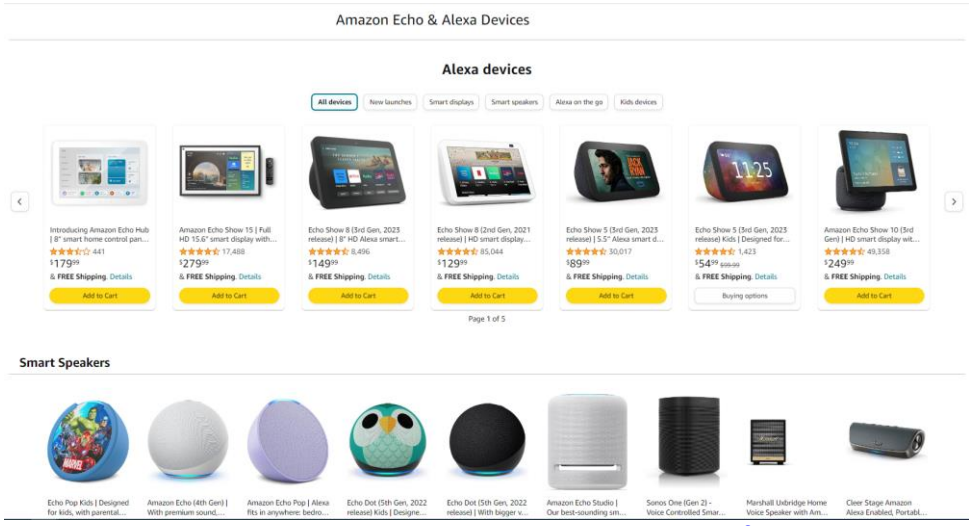
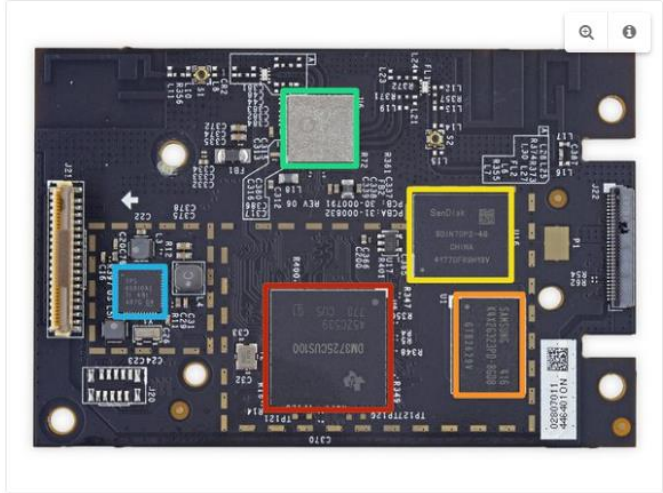
Claim	Exemplary Disclosures
	 <p>The screenshot displays the 'Alexa Privacy' interface on a mobile device. At the top, the status bar shows the time as 11:22. The app header includes a back arrow, the title 'Alexa Privacy', and a close 'X' button. Below the header, there is a section for 'Date Range' with a dropdown menu set to 'Today' and a link to 'Delete All Recordings for Today'. A separator line follows, leading to the 'Delete Selected Recordings' section. This section contains three items, each with a checkbox, a play button icon, a question, a quote, and a timestamp. The first item is 'when was the song released?' with the quote '“Hello” by Adele was released on October 23, 2015.' and a timestamp of 'Today at 11:21 AM on [redacted] Echo Dot'. The second item is 'who sings that song?' with the quote '“This is Hello from 25 by Adele.”' and the same timestamp. The third item is 'alexa play hello' with the quote '“Here’s Hello, by Adele, on Amazon Music.”' and the same timestamp. Each item has a 'Delete Recording' button with a trash icon. At the bottom of the list are 'Previous' and 'Next' navigation links.</p>

EXHIBIT H

Exhibit H: Exemplary Infringement Claim Chart for U.S. Patent 11,087,385 Claim 1

'385 Patent Claim Language	Exemplary Disclosures
<p>[1.pre] A method for providing voice commerce, the method being implemented on a computer system having one or more physical processors programmed with computer program instructions which, when executed, perform the method, the method comprising:</p>	<p>Amazon's Alexa is programmed to perform and have performed methods for providing voice commerce, the method being implemented on a computer system having one or more physical processors programmed with computer program instructions which, when executed, perform the method.</p> <p>Amazon's Alexa includes for example, Echo, Echo Dot, Echo Show, FireTV, Alexa apps, and Alexa cloud.</p>  <p>The screenshot displays the Amazon website's 'Echo & Alexa Devices' section. It features a grid of product cards for various smart displays and speakers. Each card includes an image of the device, its name, release information, customer ratings (stars), price, and an 'Add to Cart' button. Some cards also mention 'FREE Shipping'. Below the grid, there is a 'Smart Speakers' section with a row of smaller product images for different speaker models like Echo Pop, Echo Dot, and Echo Studio.</p>

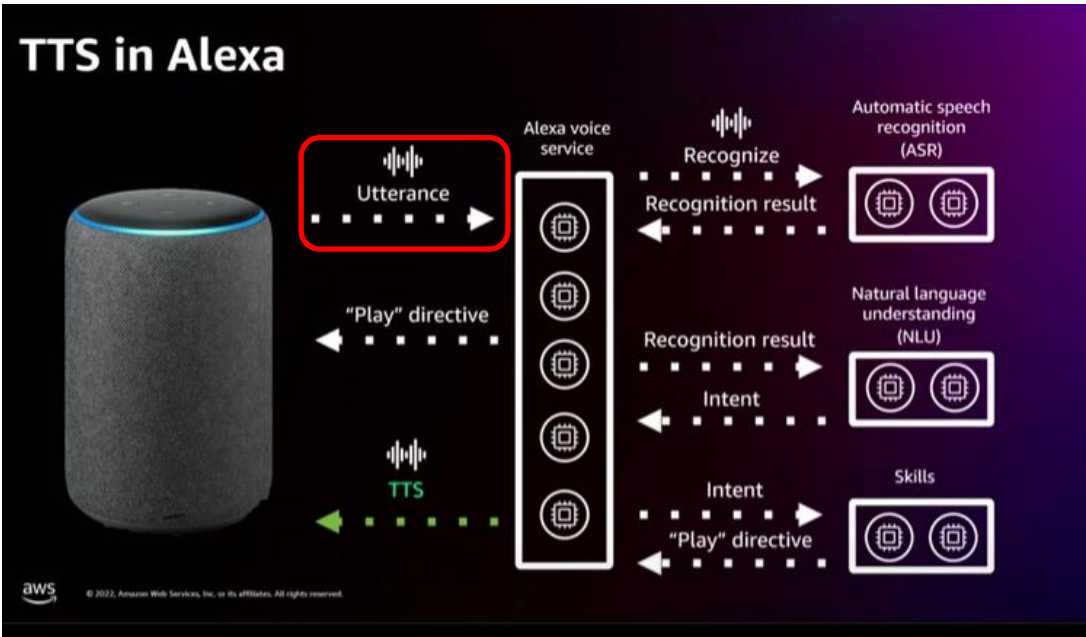
'385 Patent Claim Language	Exemplary Disclosures
	<p data-bbox="709 293 1367 313">Digital Services and Device Support › Alexa Features Help › Shopping with Alexa ›</p> <h2 data-bbox="709 318 1297 354">Turn Alexa Voice Purchasing On or Off</h2> <p data-bbox="709 383 1329 402">To turn Voice Purchasing on or off for your account, use the Alexa app.</p> <hr data-bbox="709 423 1436 427"/> <ol data-bbox="724 444 1026 561" style="list-style-type: none"> 1. Open the Alexa app . 2. Open More  and select Settings. 3. Select Account Settings. 4. Select Voice Purchasing. 5. Turn Voice Purchasing on or off. <p data-bbox="709 583 1421 625">Note: You can limit purchasing so that only recognized household members can place orders with Alexa.</p> <p data-bbox="709 649 861 669">Related Help Topics</p> <p data-bbox="709 690 1098 709">Manage an Alexa Voice ID for Purchases with Alexa</p> <div data-bbox="709 730 1194 846"> <p data-bbox="730 751 959 771">Was this information helpful?</p> <div data-bbox="730 792 835 824"> <input data-bbox="737 797 779 820" type="button" value="Yes"/> <input data-bbox="793 797 835 820" type="button" value="No"/> </div> </div> <p data-bbox="688 911 1871 943">https://www.amazon.com/gp/help/customer/display.html?nodeId=GPUCQ6PMPMENG8FA</p> <p data-bbox="688 982 1904 1083">Additionally, the voice purchasing function is implemented on Amazon's Alexa that has one or more physical processors. See https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/</p> <p data-bbox="688 1123 1887 1190">Additionally, Echo devices have one or more processors. For example, the Texas Instruments DM3725CUS100 Digital Media Processor as depicted below with red annotations:</p>

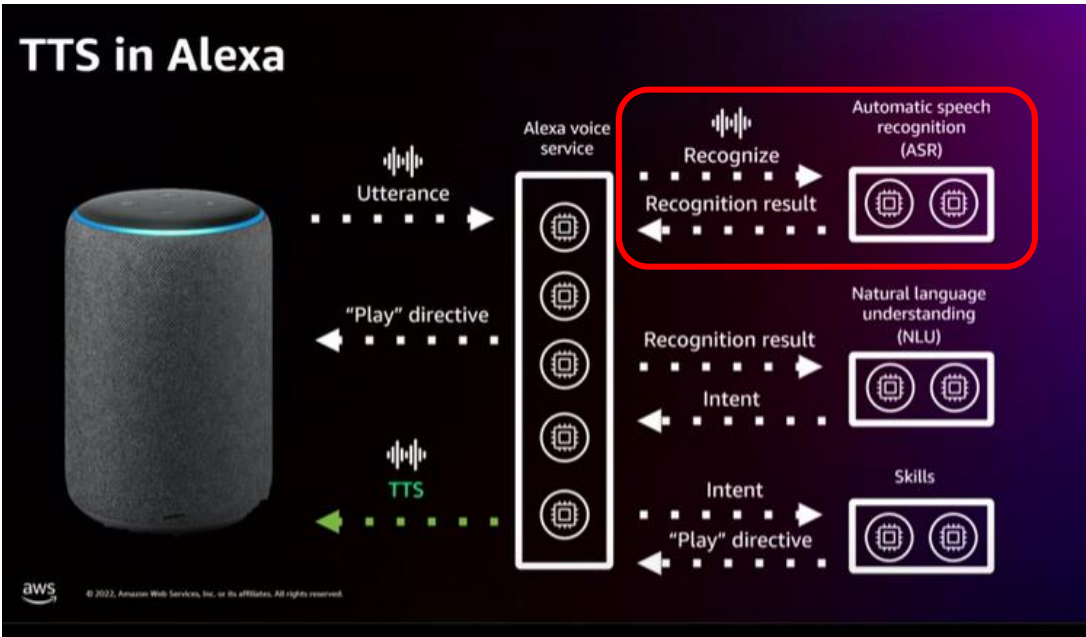
'385 Patent Claim Language	Exemplary Disclosures
	 <p data-bbox="688 764 1554 833">Sam Goldheart, Amazon Echo Teardown, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953</p>
[1.a] receiving, by the computer system, a single first user input comprising a natural language utterance;	Amazon's Alexa is programmed to perform and have performed methods for receiving, by the computer system, a single first user input comprising a natural language utterance.

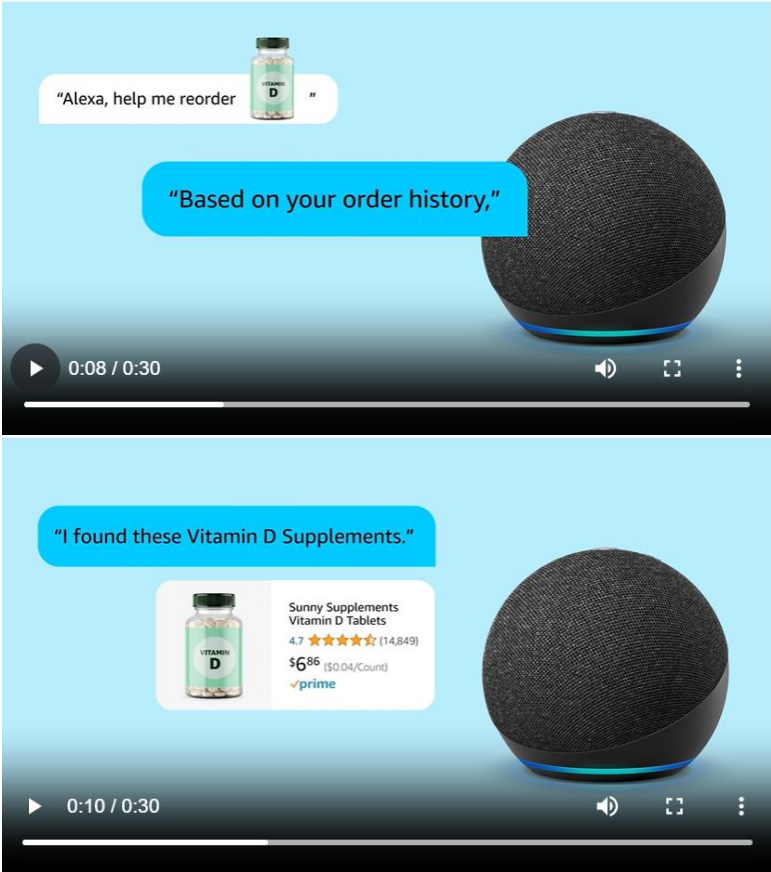
'385 Patent Claim Language	Exemplary Disclosures
	<p data-bbox="709 272 1333 292">Digital Services and Device Support › Alexa Features Help › Shopping with Alexa ›</p> <h2 data-bbox="709 297 1360 370">Manage an Alexa Voice ID for Purchases with Alexa</h2> <p data-bbox="709 399 1354 444">Use the Alexa app to turn on or turn off an Alexa voice ID to place orders with Alexa.</p> <hr data-bbox="709 462 1400 466"/> <p data-bbox="709 483 1369 524">Managing an Alexa voice ID ensures that only recognized household members place orders with Alexa.</p> <ol data-bbox="724 553 1014 711" style="list-style-type: none"> 1. Open the Alexa app . 2. Open More  and select Settings. 3. Select Account Settings. 4. Select Voice Purchasing. 5. Go to Purchase controls. 6. Select Only recognized voices. 7. Select who can make voice purchases. <p data-bbox="709 732 1392 773">Note: Digital purchases such as Kindle Books, Audible, Prime Videos, Alexa Skills, and Amazon Pay transactions, such as Alexa Donations, are excluded.</p> <p data-bbox="709 794 854 813">Related Help Topics</p> <p data-bbox="709 834 879 850">What Is Alexa Voice ID?</p> <p data-bbox="709 855 989 872">Turn Alexa Voice Purchasing On or Off</p> <div data-bbox="709 893 1169 1002"> <p data-bbox="730 912 947 928">Was this information helpful?</p> <div data-bbox="730 948 831 984"> <input data-bbox="730 948 779 984" type="button" value="Yes"/> <input data-bbox="779 948 831 984" type="button" value="No"/> </div> </div> <p data-bbox="688 1019 735 1045">See</p> <p data-bbox="688 1053 1934 1120">https://www.amazon.com/gp/help/customer/display.html?ref_=hp_left_v4_sib&nodeId=GKW8N2S4UQ2SAEBL</p> <p data-bbox="688 1159 1640 1192">User can place orders with Alexa from Amazon by uttering “order [item].”</p>

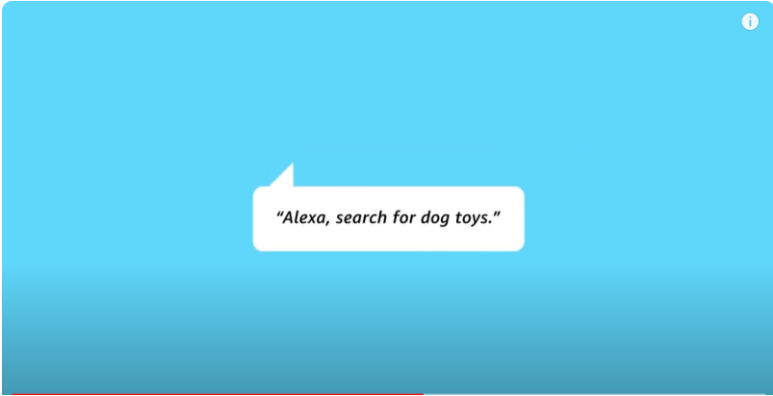
'385 Patent Claim Language	Exemplary Disclosures
	<p data-bbox="709 272 1394 293">Digital Services and Device Support › Alexa Features Help › Shopping with Alexa ›</p> <h2 data-bbox="709 300 1087 334">Place Orders with Alexa</h2> <p data-bbox="709 367 1428 415">Ask Alexa to place orders for products from your order history or your Amazon Cart.</p> <hr data-bbox="709 435 1465 440"/> <ol data-bbox="726 459 1455 557" style="list-style-type: none"> 1. Say, "Order [item]." Alexa adds a selection for that item to your Cart. Follow the prompts to place your order. Some items, require completion using the Amazon app or on the Amazon website. 2. Or, say, "Checkout my Amazon cart," and follow the prompts to check out your order. <p data-bbox="709 578 1142 599">If you no longer want the order, say "Cancel my order."</p> <div data-bbox="709 617 1213 735"> <p data-bbox="730 638 968 659">Was this information helpful?</p> <div data-bbox="730 678 842 716"> <input data-bbox="741 686 779 711" type="button" value="Yes"/> <input data-bbox="800 686 837 711" type="button" value="No"/> </div> </div> <p data-bbox="688 760 735 789">See</p> <p data-bbox="688 795 1944 860">https://www.amazon.com/gp/help/customer/display.html?ref=hp_left_v4_sib&nodeId=GFEBUXE7DH7G9EQ4</p> <p data-bbox="722 889 1638 1003">Say something like, "Order (item) from Prime Now." If the item is available for two-hour delivery, it will be added to your cart and you'll need to order and confirm the item to have it shipped promptly.</p> <p data-bbox="722 1060 1608 1174">If you know exactly what you want, tell Alexa. Mention brand, quantity and color. For example, "Alexa, add two black Sharpie markers to my shopping cart."</p> <p data-bbox="688 1185 1892 1250">See https://www.cnet.com/home/smart-home/alexa-shopping-and-voice-purchasing-a-step-by-step-guide/</p> <p data-bbox="688 1291 1892 1356">Users can ask Alexa to reorder a past order. For example, one can say "Alexa, help me reorder Vitamin D."</p>

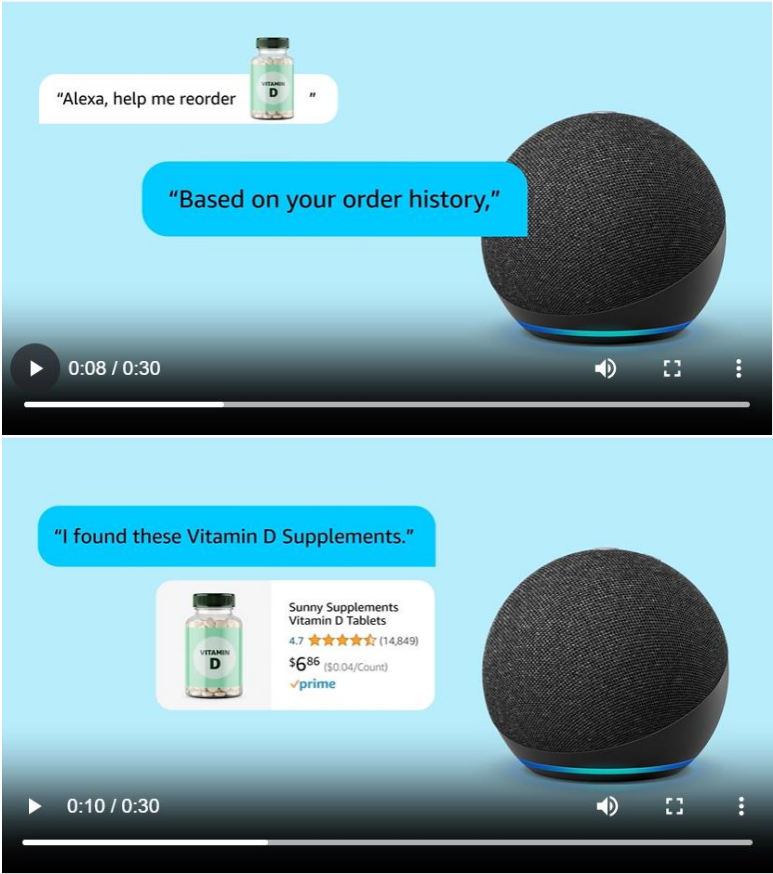
'385 Patent Claim Language	Exemplary Disclosures
	 <p>https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011&ref=pe_alxhub_aucc_en_us_IC_HP_3_HUB_SHOP</p> <p>Users can also ask Alexa to place an order on an item that hasn't been purchased before. E.g., "Alexa, order triple A batteries" to place an order of triple A batteries.</p> <p>https://www.youtube.com/watch?v=M_8-xJA668c</p>
[1.b] providing, by the computer system, the natural language utterance as an input to a speech recognition engine;	<p>Alexa is programmed to perform and have performed methods for providing, by the computer system, the natural language utterance as an input to a speech recognition engine.</p> <p>Alexa has a speech recognition engine that takes user's natural language utterance as inputs.</p>

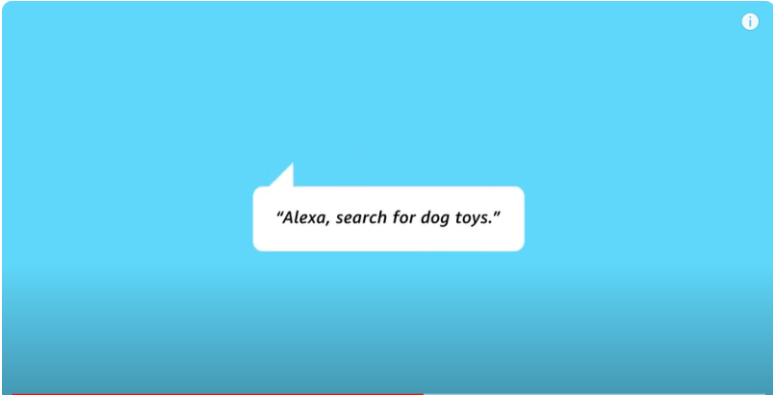
'385 Patent Claim Language	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>The diagram illustrates the TTS in Alexa process. It shows an Alexa device on the left. A red box highlights the 'Utterance' input to the 'Alexa voice service' (represented by a vertical stack of four processor icons). The process flows from the Alexa voice service to three components: 'Automatic speech recognition (ASR)', 'Natural language understanding (NLU)', and 'Skills'. Each component sends a 'Recognition result' back to the Alexa voice service. The 'NLU' component sends an 'Intent' back to the Alexa voice service. The 'Skills' component sends a '"Play" directive' back to the Alexa voice service. The Alexa voice service then outputs a '"Play" directive' to the Alexa device and a 'TTS' (Text-to-Speech) output to the Alexa device. The AWS logo and copyright notice are at the bottom left of the diagram.</p> <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y, at 10:07.</p>
<p>[1.c] obtaining, by the computer system, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine;</p>	<p>Alexa is programmed to perform and have performed methods for obtaining, by the computer system, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.</p>

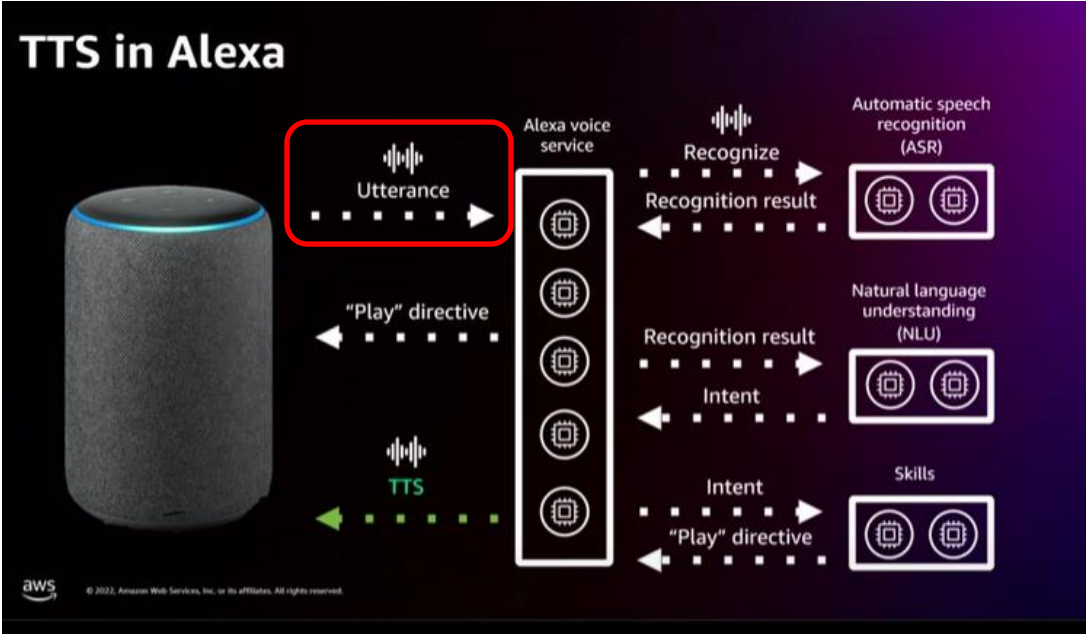
'385 Patent Claim Language	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>The diagram illustrates the TTS in Alexa process. It shows an Alexa device (a smart speaker) interacting with the Alexa voice service. The process involves three main components: Automatic speech recognition (ASR), Natural language understanding (NLU), and Skills. The ASR component receives an 'Utterance' and returns a 'Recognition result'. The NLU component receives a 'Recognition result' and returns an 'Intent'. The Skills component receives an 'Intent' and returns a '"Play" directive'. The TTS component then takes the '"Play" directive' and outputs the final audio response. The ASR component is highlighted with a red box.</p> <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y, at 10:07.</p>
<p>[1.d] searching, by the computer system, one or more databases of products or services based on the one or more words or phrases;</p>	<p>Alexa is programmed to perform and have performed methods of searching, by the computer system, one or more databases of products or services based on the one or more words or phrases.</p> <p>If the user has purchased the same product before, Alexa will search the user's existing order history.</p>

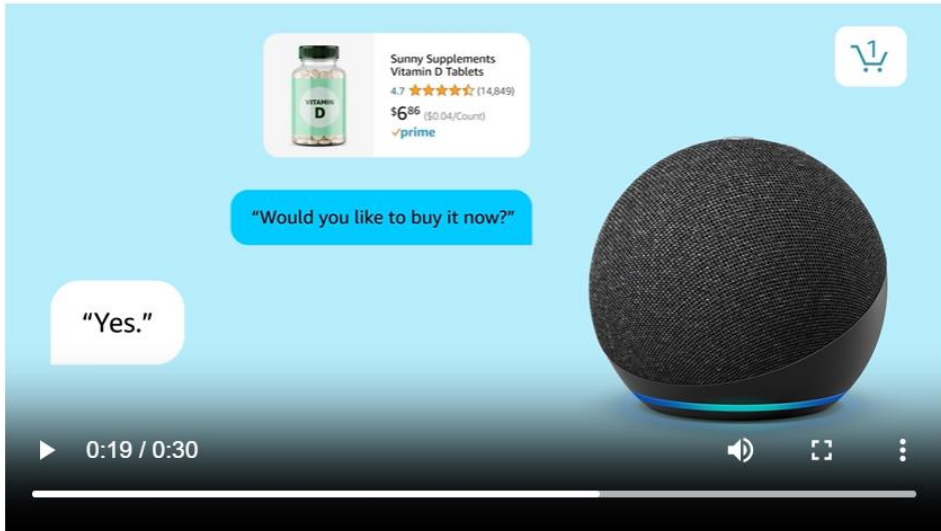
'385 Patent Claim Language	Exemplary Disclosures
	 <p>See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011&ref=pe_alxhub_aucc_en_us_IC_HP_3_HUB_SHOP</p> <p>If the user has not purchased the same product before, Alexa will search the entire Amazon database.</p>

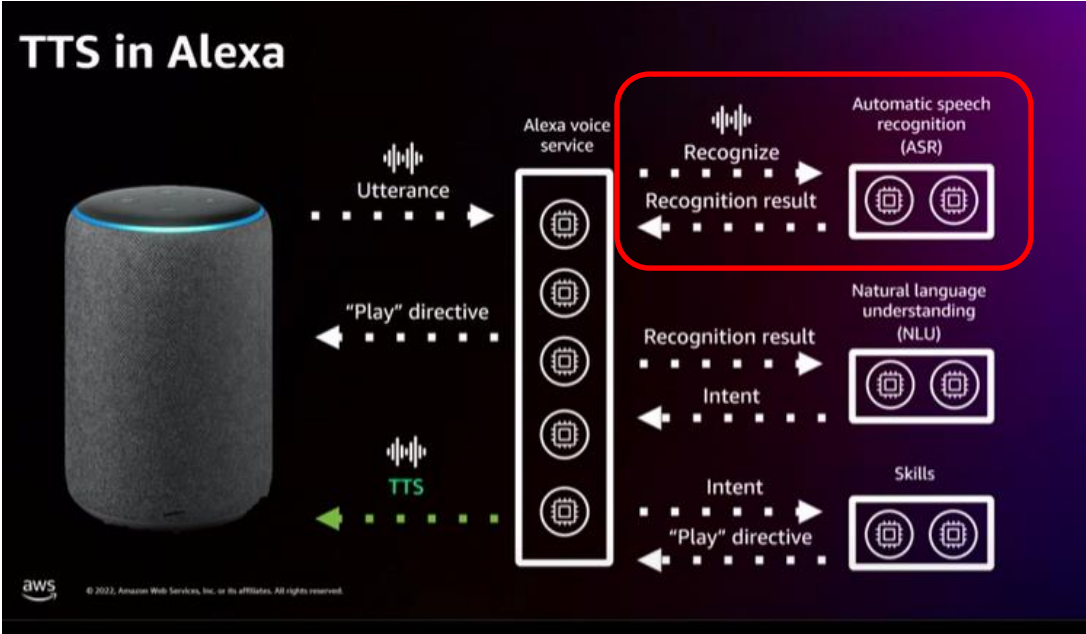
'385 Patent Claim Language	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=kEmo9CmFshQ , at 0:19</p> <p>Alexa searches the database for Amazon's alkaline batteries. See https://www.youtube.com/watch?v=M_8-xJA668c</p>
<p>[1.e] selecting, by the computer system, without further user input other than the single first user input, a product or service from the database to be purchased based on the search;</p>	<p>Alexa is programmed to perform and have performed methods of selecting, by the computer system, without further user input other than the single first user input, a product or service from the database to be purchased based on the search.</p> <p>If the user has purchased the same product before, Alexa will select the product based on order history.</p>

'385 Patent Claim Language	Exemplary Disclosures
	 <p>See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011&ref=pe_alxhub_aucc_en_us_IC_HP_3_HUB_SHOP</p> <p>If the user has not purchased the same product before, Alexa will search the entire Amazon database and select Amazon's choice product.</p>

'385 Patent Claim Language	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=kEmo9CmFshQ , at 0:19</p> <p>Alexa selects Amazon's Choice—"Amazon's alkaline batteries triple A 36 pack."</p> <p>See https://www.youtube.com/watch?v=M_8-xJA668c</p>
<p>[1.f] receiving, by the computer system, a second user input indicating confirmation by a user to complete a purchase transaction of the selected product or service; and</p>	<p>Alexa is programmed to perform and have performed methods for receiving, by the computer system, a second user input indicating confirmation by a user to complete a purchase transaction of the selected product or service.</p>

'385 Patent Claim Language	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y, at 10:07.</p> <p>The user will have to say yes before completing a purchase transaction of the selected product or service.</p>

'385 Patent Claim Language	Exemplary Disclosures
	 <p>See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011&ref=pe_alxhub_aucc_en_us_IC_HP_3_HUB_SHOP</p> <p>After Alexa selects “Amazon’s alkaline batteries triple A 36 pack,” Alexa asks the user “would you like to buy it?” The user then says “yes” as a second user input indicating confirmation by a user to complete the purchase transaction of triple A batteries. See https://www.youtube.com/watch?v=M_8-xJA668c</p>
[1.g] completing, by the computer system, without further user input after the receipt of the second user input, a purchase transaction of the selected product or service.	Alexa is programmed to perform and have performed methods for receiving, by the computer system, a second user input indicating confirmation by a user to complete a purchase transaction of the selected product or service.

'385 Patent Claim Language	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>See https://www.youtube.com/watch?v=gRAEp--sZ_Y, at 10:07.</p> <p>A purchase transaction is automatically completed after receiving the second user input.</p>

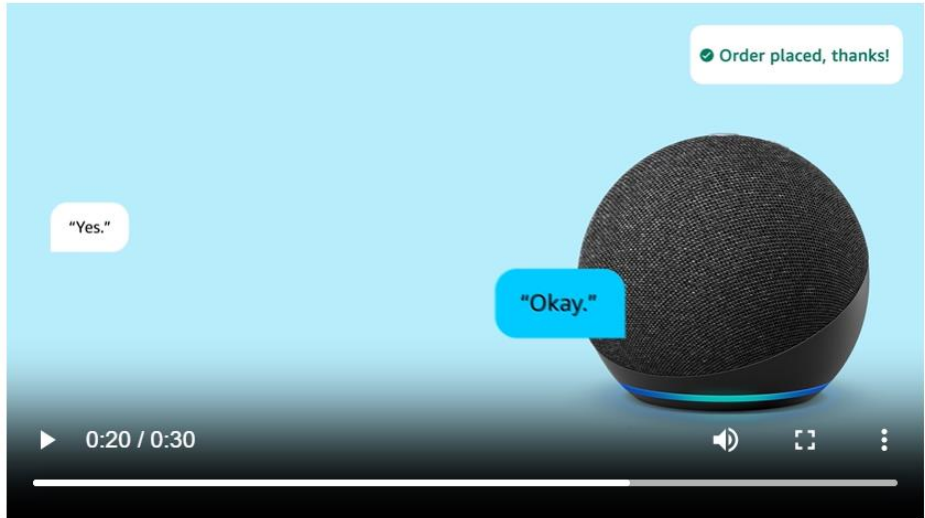
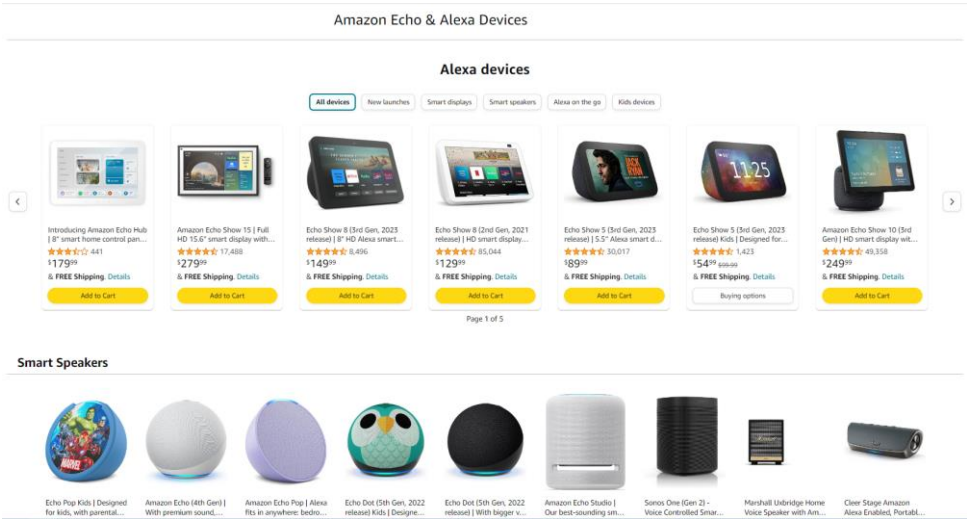
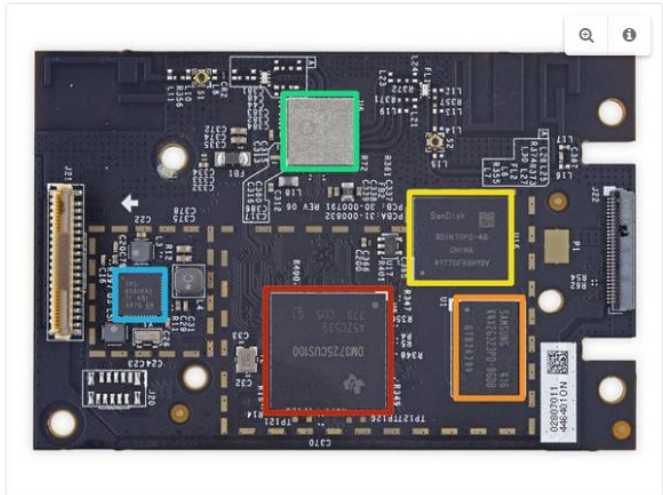
'385 Patent Claim Language	Exemplary Disclosures
	<div data-bbox="690 261 1608 776">A screenshot of an Amazon Echo device interface. The device is a black, spherical smart speaker with a blue light ring at the base. The screen displays a confirmation message 'Order placed, thanks!' in a green bubble. Below it, there is a white bubble with the text 'Yes.' and a blue bubble with the text 'Okay.' The video player controls at the bottom show a progress bar at 0:20 / 0:30.</div> <p data-bbox="690 787 1864 852">See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011&ref=pe_alxhub_aucc_en_us_IC_HP_3_HUB_SHOP</p> <p data-bbox="690 893 1885 958">After the user then says “yes” as a second user input indicating confirmation, Alexa places the order and says “thanks, order placed.” See https://www.youtube.com/watch?v=M_8-xJA668c</p>

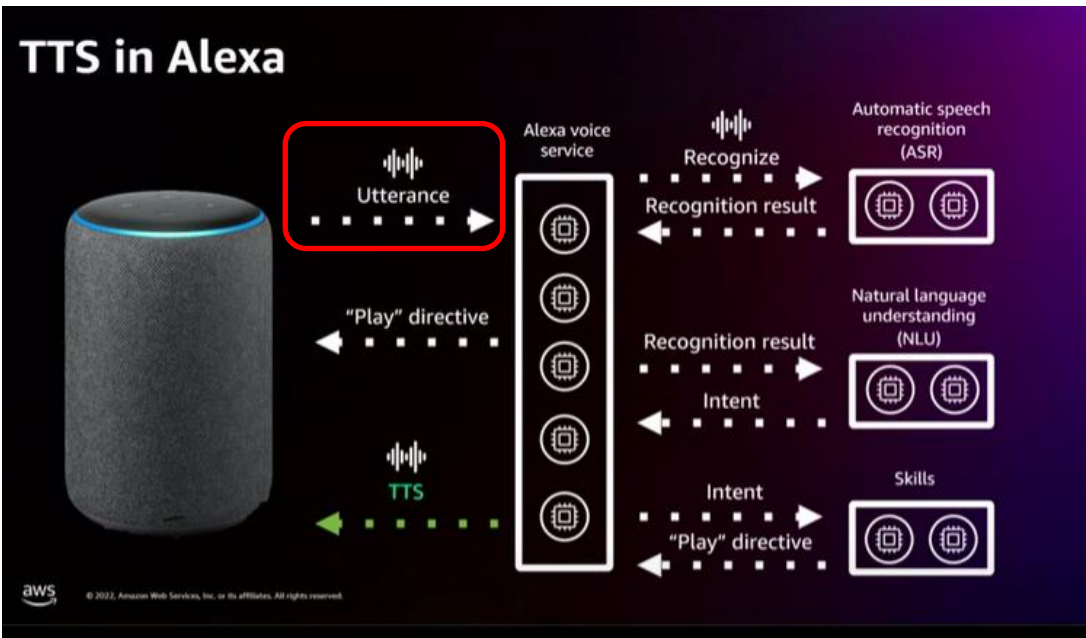
EXHIBIT I

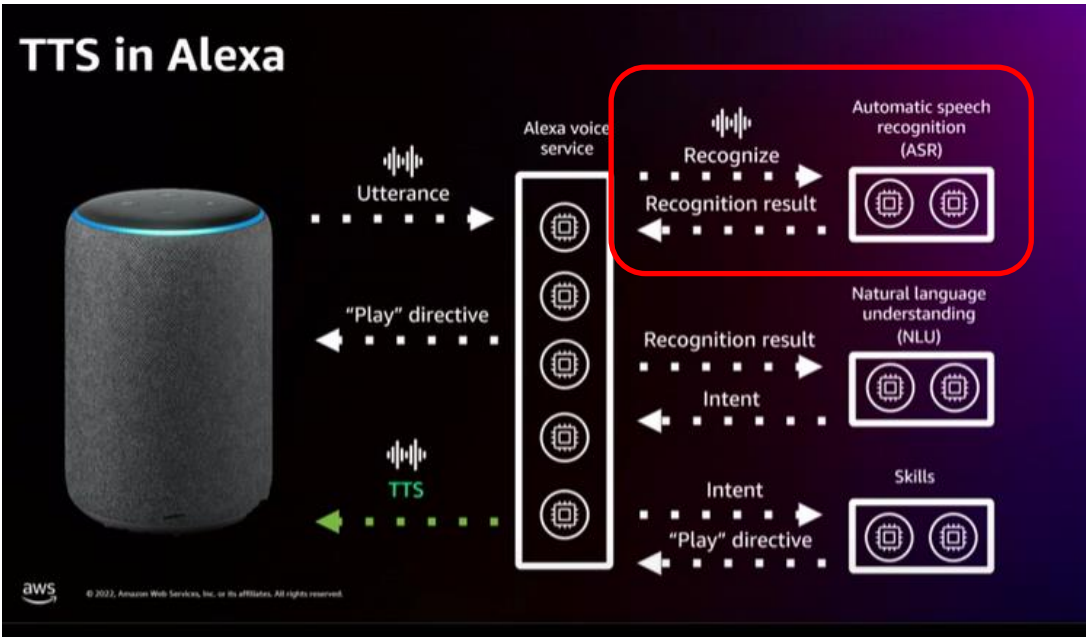
Exhibit I: Exemplary Infringement Claim Chart for U.S. Patent 11,080,758 Claim 1

Claim	Exemplary Disclosures
<p>[1.pre] A method for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon, the method being implemented by one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method, the method comprising:</p>	<p>Amazon's Alexa is programmed to perform and have performed methods for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon, the method being implemented by one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method.</p> <p>Amazon's Alexa includes for example, Echo, Echo Dot, Echo Show, FireTV, Alexa apps, and Alexa cloud.</p>  <p>The screenshot shows the Amazon Echo & Alexa Devices product page. It features a grid of various smart display devices including Echo Show 15, Echo Show 10, Echo Show 8, Echo Show 5, and Echo Show 10 (3rd Gen). Below the grid, there is a section for Smart Speakers featuring Echo Pop Kids, Amazon Echo (4th Gen), Amazon Echo Flex, Echo Dot (5th Gen), Echo Dot (3rd Gen), Amazon Echo Studio, Sonos One, Marshall Midrange Home Voice Speaker, and Clear Stage Amazon Alexa Enabled. Each product listing includes an image, a brief description, a star rating, a price, and an 'Add to Cart' button.</p> <p>See https://www.amazon.com/b?&node=9818047011&ref=MARS_NAV_desktop_plswitcher</p> <p>Alexa users can use the voice purchasing function.</p>

Claim	Exemplary Disclosures
	<p data-bbox="709 293 1367 315">Digital Services and Device Support › Alexa Features Help › Shopping with Alexa ›</p> <h2 data-bbox="709 321 1297 358">Turn Alexa Voice Purchasing On or Off</h2> <p data-bbox="709 386 1325 407">To turn Voice Purchasing on or off for your account, use the Alexa app.</p> <hr data-bbox="709 423 1436 428"/> <ol data-bbox="724 444 1026 565" style="list-style-type: none"> 1. Open the Alexa app . 2. Open More  and select Settings. 3. Select Account Settings. 4. Select Voice Purchasing. 5. Turn Voice Purchasing on or off. <p data-bbox="709 586 1419 630">Note: You can limit purchasing so that only recognized household members can place orders with Alexa.</p> <p data-bbox="709 651 863 672">Related Help Topics</p> <p data-bbox="709 693 1100 714">Manage an Alexa Voice ID for Purchases with Alexa</p> <div data-bbox="709 732 1192 846"> <p>Was this information helpful?</p> <div> <input data-bbox="737 797 779 818" type="button" value="Yes"/> <input data-bbox="800 797 842 818" type="button" value="No"/> </div> </div> <p data-bbox="688 915 1871 948">https://www.amazon.com/gp/help/customer/display.html?nodeId=GPUCQ6PMPMENG8FA</p> <p data-bbox="688 987 1902 1094">Additionally, the voice purchasing function is implemented on Amazon's Alexa that has one or more physical processors. See https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/</p> <p data-bbox="688 1133 1885 1203">Additionally, Echo devices have one or more processors. For example, the Texas Instruments DM3725CUS100 Digital Media Processor as depicted below with red annotations:</p>

Claim	Exemplary Disclosures
	 <p data-bbox="688 764 1554 836">Sam Goldheart, Amazon Echo Teardown, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953</p>
<p data-bbox="218 911 642 1052">[1.a] providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;</p>	<p data-bbox="688 911 1871 979">Alexa programmed to perform or have performed methods for providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine.</p>

Claim	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y, at 10:07.</p>
<p>[1.b] receiving, by the one or more processors, words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine;</p>	<p>Alexa is programmed to perform and have performed methods for receiving, by the one or more processors, words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine.</p>

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y, at 10:07.</p>
<p>[1.c] determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;</p>	<p>Alexa is programmed to perform or have performed methods of determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases.</p> <p>“Recently, Alexa also introduced a context embedding service, which uses a large neural network trained on a variety of tasks to produce a running sequence of vector representations — or embeddings — of the past several rounds of dialogue, both the customer’s utterances and Alexa’s responses.”</p> <p>“The underlying machine learning model factors in the current customer utterance, the text of the previous dialogue turn (both the customer’s utterance and Alexa’s response), and relevant context information from the Alexa services invoked by the utterance. This might include entries from an address book, a list of smart-home devices connected to Alexa, or the local-search function’s</p>

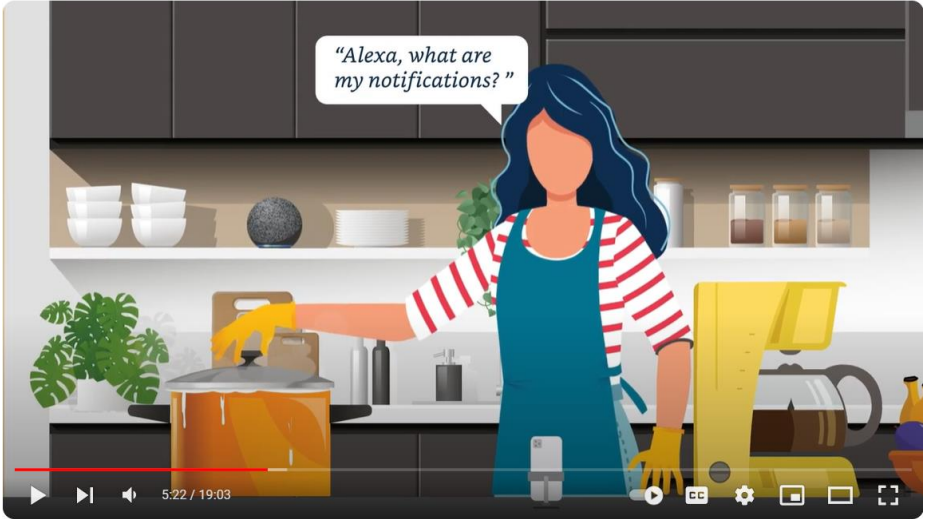
Claim	Exemplary Disclosures										
	<p>classification of names the customer mentioned — names of restaurants, of movie theaters, of gas stations, and so on.”</p> <p>https://www.amazon.science/latest-news/the-engineering-behind-alexa-contextual-speech-recognition</p> <h3 data-bbox="821 509 1142 548">Contextual carryover</h3> <p data-bbox="821 586 1818 727">You don't need to provide annotated dialogs that demonstrate contextual carryover. The dialog simulator supports this feature by introducing turns with pronouns (for example, "Please purchase it") when generating dialog variants from annotated dialogs. The carryover occurs at a later runtime stage, argument filling, when Alexa Conversations considers all slots that the user and Alexa mention across the entire dialog for filling API arguments.</p> <p data-bbox="821 756 903 779">Example</p> <p data-bbox="821 812 1722 834">You provide a dialog to purchase a movie. The dialog simulator introduces a turn with a pronoun.</p> <table data-bbox="821 865 1818 1110"> <thead> <tr> <th data-bbox="829 872 1297 914">Dialog you provide</th><th data-bbox="1318 872 1810 914">Example variant from the dialog simulator</th></tr> </thead> <tbody> <tr> <td data-bbox="829 927 1297 982">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td><td data-bbox="1318 927 1810 982">User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)</td></tr> <tr> <td data-bbox="829 985 1297 1008">Alexa: How about Snatch?</td><td data-bbox="1318 985 1810 1008">Alexa: How about Snatch?</td></tr> <tr> <td data-bbox="829 1011 1297 1066">User: Please help me purchase Snatch. (Invoke API ReserveMovie.)</td><td data-bbox="1318 1011 1810 1066">User: Please purchase it. (Invoke API ReserveMovie.)</td></tr> <tr> <td data-bbox="829 1070 1297 1092">Alexa: The movie Snatch is reserved for you.</td><td data-bbox="1318 1070 1810 1092">Alexa: The movie Snatch is reserved for you.</td></tr> </tbody> </table> <h3 data-bbox="821 1149 1079 1188">Context memory</h3> <p data-bbox="821 1226 1810 1339"><i>Context memory</i> is short-term memory that tracks user utterances and models results such as entities (that is, a slot that the user mentioned or a return value from a previous API), predicted actions, and Alexa responses. Context memory maintains dialog context for each session and is constantly updated throughout the session.</p> <p data-bbox="688 1365 1881 1437">See https://developer.amazon.com/en-US/docs/alexa/conversations/how-alexa-conversations-works.html#contextual-carryover</p>	Dialog you provide	Example variant from the dialog simulator	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	Alexa: How about Snatch?	Alexa: How about Snatch?	User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)	Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.
Dialog you provide	Example variant from the dialog simulator										
User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)	User: I'd like to watch a movie from Guy Ritchie. (Invoke API RecommendMovie.)										
Alexa: How about Snatch?	Alexa: How about Snatch?										
User: Please help me purchase Snatch. (Invoke API ReserveMovie.)	User: Please purchase it. (Invoke API ReserveMovie.)										
Alexa: The movie Snatch is reserved for you.	Alexa: The movie Snatch is reserved for you.										

Claim	Exemplary Disclosures
	<p data-bbox="814 316 1570 354">What do events and directives look like?</p> <div data-bbox="751 386 1186 763"><pre data-bbox="762 394 1176 711">SpeechRecognizer.Recognize 0027-10-10 10:10:25 { "event": { "header": { "dialogRequestId": "dialogRequestId-2fa1867c-bfe2-4113-9bd-325a1c466156", "namespace": "SpeechRecognizer", "name": "recognize", "messageId": "e4717a2e-160b-4153-ae31-3068107542be" }, "payload": { "profile": "CLOSE_TALK", "format": "AUDIO_L16_RATE_16000_CHANNELS_1" } }, "context": { { "header": { "namespace": "AudioPlayer", "name": "PlaybackState" }, "payload": { "token": "", "playerActivity": "IDLE", "offsetInMilliseconds": "0" } } } }</pre></div> <p data-bbox="926 768 995 792">Event</p> <div data-bbox="1224 386 1507 500"><p data-bbox="1308 402 1507 427">Alexa, what time is it?</p></div> <p data-bbox="730 808 852 849">AWS re:Invent</p> <p data-bbox="730 841 1045 849">© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.</p> <p data-bbox="1602 800 1661 841">aws</p> <p data-bbox="688 901 1927 933">AWS re: Invent, Integrate Alexa into Your Product, Using the Alexa Voice Service Device SDK.</p>

Claim	Exemplary Disclosures
	<div data-bbox="688 256 1703 846"> <h3>Conversation as context</h3> <p>Conversation Context</p> <p>“How are you?”</p> <p>“I’m doing fantastic today!”</p> <p>“I’m having an awful day!”</p> <p>“What happened?”</p> <p>Text</p> <p>Emotional prosody space</p> <ul style="list-style-type: none"> - Valence - Arousal - Dominance <p>TTS</p> <p>aws © 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.</p> </div> <div data-bbox="688 846 1703 1414"> <h3>Adapting speech to the conversation</h3> <p>Alexa, help me find some music.</p> <p>Happy to help. Let's find something you'll like on Amazon Music. Any particular type of music you wanted?</p> <p>Let's have some rock music.</p> <p>Got it. How about something by The Eagles?</p> <p>I'm not a big fan of The Eagles.</p> <p>Hmm, okay. How about we try some songs by Queen?</p> <p>No, give me something more modern!</p> <p>Oh dear, I'm not getting you at all today. What would you think about something by Foo Fighters?</p> <p>That's more like it.</p> <p>Great. In that case, do you think you'd like Vampire Weekend?</p> <p>Now you're getting it . . .</p> <p>Fantastic. Pulling a playlist together for you right now.</p> <p>aws © 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.</p> </div>

See https://www.youtube.com/watch?v=gRAFP--sZ_Y

Claim	Exemplary Disclosures
<p>[1.d] selecting, by the one or more processors, a purchase opportunity based on the determined context;</p>	<p>Alexa is programmed to perform or have performed methods of selecting, by the one or more processors, a purchase opportunity based on the determined context. Alexa selects a purchase opportunity, e.g. an artist's upcoming tour ticket, deals, etc. based on the determined context.</p> <p>“What are interest-based ads? Interest-based ads are used to highlight products and services that may be relevant to you based on your interests. For example, if you browse cake pans on Amazon.com, you may later receive a relevant ad on your Alexa-enabled device about cookware. Or, if you often ask Alexa to play a song from your favorite artist when listening to a third-party streaming service, you may get an ad when listening to that service promoting that artist's upcoming tour.” https://www.amazon.com/b?ie=UTF8&node=98592480011#:~:text=There%20are%20experience%20with%20Alexa,Show%20screen%2C%20that%20include%20ads.</p> <p style="text-align: center;">Alexa can help you save money</p> <div style="display: flex; align-items: center;">  <div style="margin-left: 20px;"> <p>Alexa can notify you about live deals on items that you have saved in your Cart, Wish List, or Saved for Later. Here's how it works:</p> <ol style="list-style-type: none"> 1. Add products to your Cart, Wish List, or Saved for Later on your Alexa device or the Amazon app. 2. When you see a yellow ring light or a notification pop-up on your Echo Device or Amazon app, ask "Alexa, what are my notifications?" Alexa will then notify you when added products go live on deal. 3. Ready to save? You can then ask Alexa to buy the deal. <p>Or, discover deals with your Echo device.</p> </div> </div> <p>See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011</p>

Claim	Exemplary Disclosures
	<p>Browsing for a gift or a product for yourself?</p> <p>Alexa can give recommendations based on real customer ratings and let you know of great deals on Amazon.com.</p> <p>Ask: "What's the most popular dog food?" Ask: "What are my deals?"</p>  <p>See https://www.amazon.com/alexa-voice-shopping/b?ie=UTF8&node=14552177011</p>  <p>https://www.youtube.com/watch?v=w73EMk_qLI0</p>
[1.e] delivering, by the one or more processors, the selected purchase opportunity via an electronic device in	<p>Alexa is programmed to perform or have performed methods of delivering, by the one or more processors, the selected purchase opportunity via an electronic device in communication with the one or more processors.</p> <p>Alexa delivers the selected purchase opportunity via Alexa devices.</p>

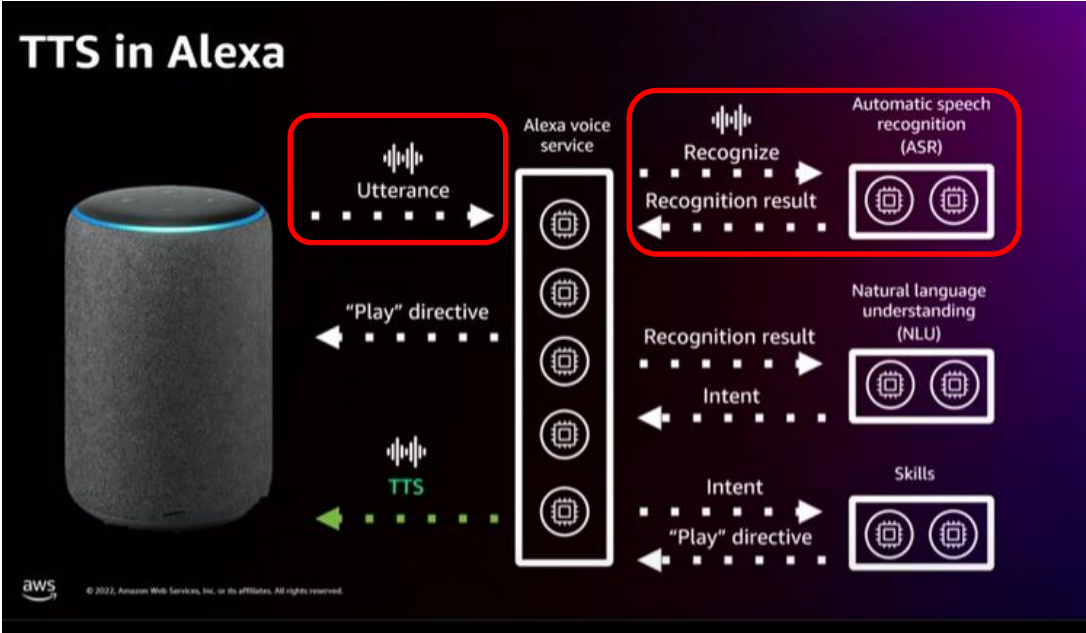
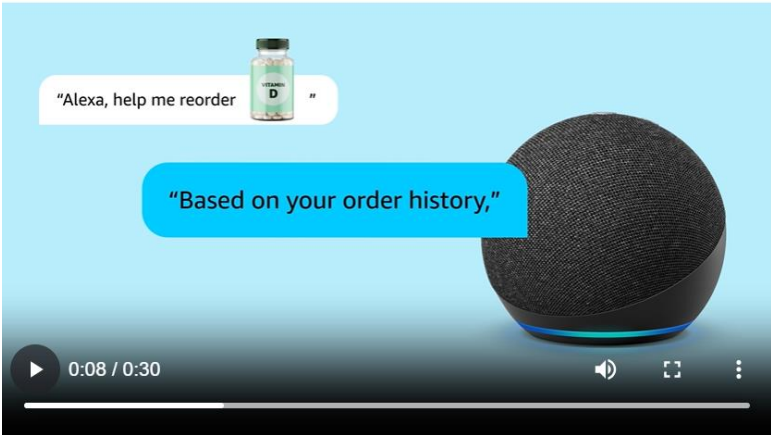
Claim	Exemplary Disclosures
communication with the one or more processors;	<p>“What are interest-based ads? Interest-based ads are used to highlight products and services that may be relevant to you based on your interests. For example, if you browse cake pans on Amazon.com, you may later receive a relevant ad on your Alexa-enabled device about cookware. Or, if you often ask Alexa to play a song from your favorite artist when listening to a third-party streaming service, you may get an ad when listening to that service promoting that artist’s upcoming tour.”</p> <p>https://www.amazon.com/b?ie=UTF8&node=98592480011#:~:text=There%20are%20experiences%20with%20Alexa,Show%20screen%2C%20that%20include%20ads</p> <p style="text-align: center;">Alexa can help you save money</p> <div style="display: flex; align-items: center;">  <div style="margin-left: 20px;"> <p>Alexa can notify you about live deals on items that you have saved in your Cart, Wish List, or Saved for Later. Here's how it works:</p> <ol style="list-style-type: none"> 1. Add products to your Cart, Wish List, or Saved for Later on your Alexa device or the Amazon app. 2. When you see a yellow ring light or a notification pop-up on your Echo Device or Amazon app, ask "Alexa, what are my notifications?" Alexa will then notify you when added products go live on deal. 3. Ready to save? You can then ask Alexa to buy the deal. <p>Or, discover deals with your Echo device.</p> </div> </div> <p>See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011</p>

Claim	Exemplary Disclosures
	<p data-bbox="716 285 961 329">Browsing for a gift or a product for yourself?</p> <p data-bbox="716 342 974 396">Alexa can give recommendations based on real customer ratings and let you know of great deals on Amazon.com.</p> <p data-bbox="716 415 991 433">Ask: "What's the most popular dog food?"</p> <p data-bbox="716 438 888 456">Ask: "What are my deals?"</p>  <p data-bbox="688 618 1785 651">See https://www.amazon.com/alexa-voice-shopping/b?ie=UTF8&node=14552177011</p> 

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=w73EMk_qLl0</p>
<p>[1.f] tracking, by the one or more processors, an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity;</p>	<p>Alexa is programmed to perform or have performed methods of tracking, by the one or more processors, an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity.</p>

Claim	Exemplary Disclosures
	<div data-bbox="688 256 1612 776">  <p>The image shows a woman with long dark hair, wearing a red and white striped shirt and a blue apron, standing in a kitchen. She is wearing yellow gloves and is reaching for a jar on a shelf. A speech bubble above her head says "yes". The kitchen has dark cabinets, a white countertop, and various kitchen items like a potted plant, a coffee maker, and jars.</p> </div> <p data-bbox="688 782 1402 815">See https://www.youtube.com/watch?v=w73EMk_qLl0</p> <p data-bbox="701 841 1390 863"> Digital Services and Device Support › Alexa Features Help › Alexa Smart Home › </p> <h3 data-bbox="701 870 1423 951">What Are Low Supply Notifications & Smart Reorders?</h3> <p data-bbox="701 984 1381 1006">Alexa can notify you or reorder essential items when you're running low.</p> <hr data-bbox="701 1029 1482 1032"/> <p data-bbox="701 1045 1440 1091">After connecting a device with Alexa that is compatible with smart reorders, like a printer, Alexa activates low supply notifications automatically.</p> <p data-bbox="701 1110 1461 1156">With low supply notifications, you'll receive notifications for compatible products when your supply is running low.</p> <p data-bbox="701 1175 1465 1221">Alternatively, activate smart reorders for automatic reordering on products you choose when Alexa detects low supply.</p> <p data-bbox="701 1243 1449 1318">Note: Low supply notifications and smart reorders can't be active at the same time. If you activate smart reorders, it turns off low supply notifications. And if you activate low supply notifications, it turns off smart reorders.</p> <p data-bbox="688 1341 1848 1373">See https://www.amazon.com/gp/help/customer/display.html?nodeId=TseBdhTjISJofpzfl0</p>

Claim	Exemplary Disclosures
<p>[1.g] building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and</p>	<p>Alexa is programmed to perform or have performed methods of building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern.</p> <p>“You can enroll your Amazon Profile in Alexa as an Alexa Profile. Alexa Profiles gives each person in your home their own individual Alexa experience. You can create an Alexa Profile for you and your family members when setting up your device. Alexa Profiles can be created from your device after setup. Select an existing Amazon Profile on your Amazon Account or create a new Amazon Profile. Existing Amazon Profiles include profiles enrolled in other programs, like Prime Video.”</p> <p>https://www.amazon.com/gp/help/customer/display.html?nodeId=GLE4EVA2VLDRQDLC</p> <p>Alexa tracks purchases for a given user. Alexa provides recommendations based on past purchase patterns. For example, if the end-user utters: “I want to reorder something,” Alexa will focus on recommending previously purchased items for that specific user.</p>
<p>[1.h] interpreting, by the one or more processors, a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.</p>	<p>Alexa is programmed to perform or have performed methods of interpreting, by the one or more processors, a subsequent natural language utterance using the user-specific profile (e.g., Amazon profile or VoiceID) to select a subsequent purchase opportunity based on the subsequent natural language utterance.</p>

Claim	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>See https://www.youtube.com/watch?v=gRAFp--sZ_Y, at 10:07.</p> 

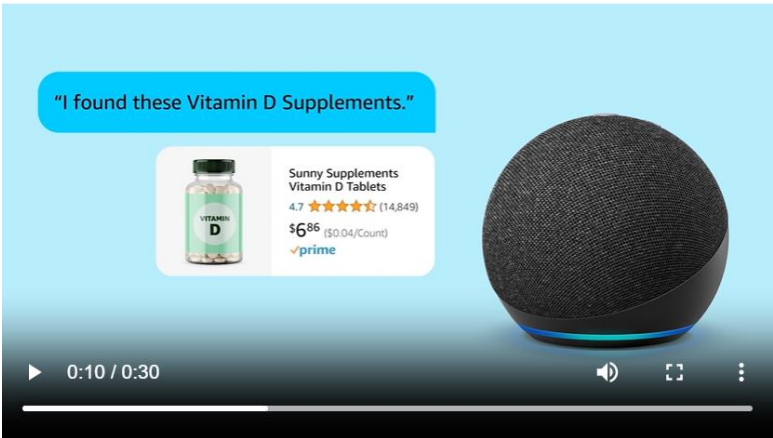
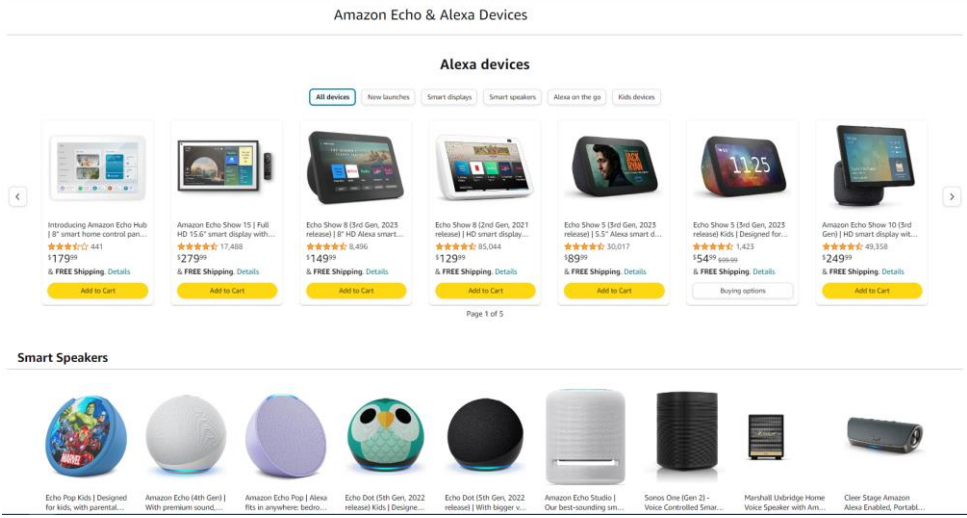
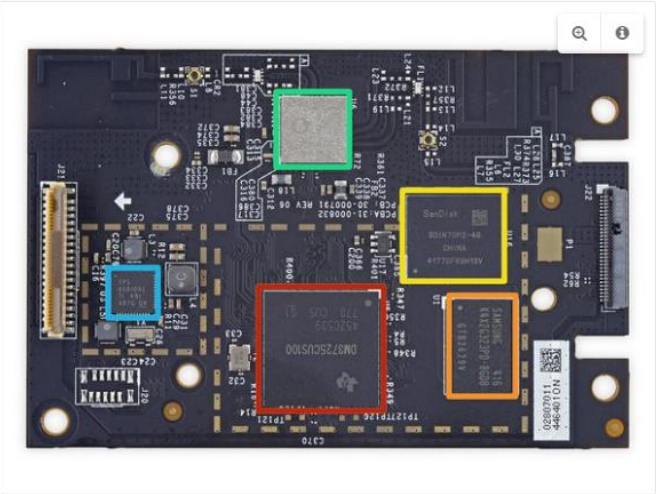
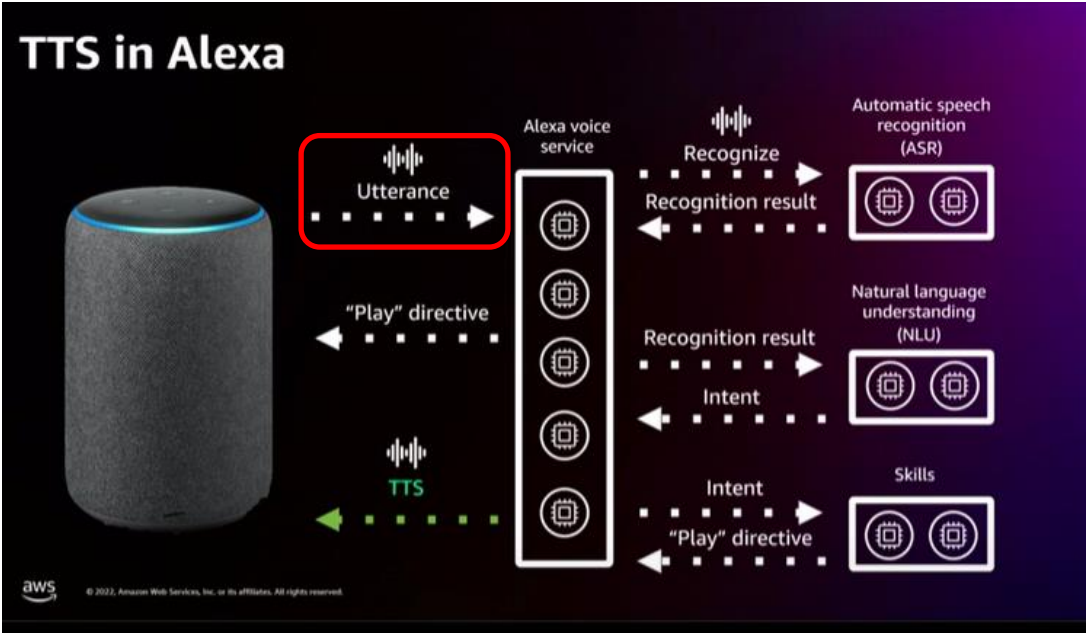
Claim	Exemplary Disclosures
	 <p>See https://www.amazon.com/alexa-shopping-hub/b?ie=UTF8&node=21467932011&ref=pe_alxhub_aucc_en_us_IC_HP_3_HUB_SHOP</p>

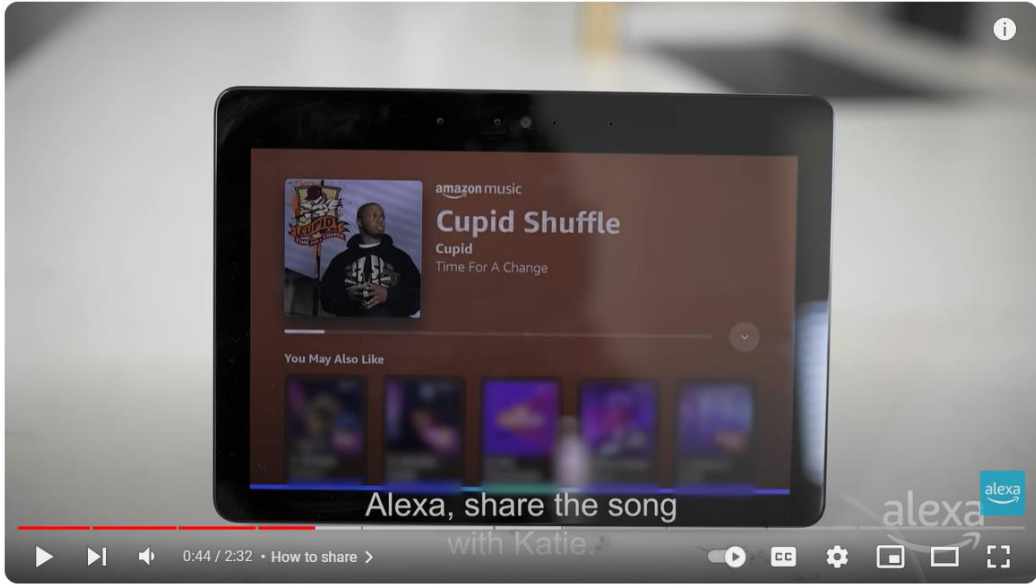
EXHIBIT J

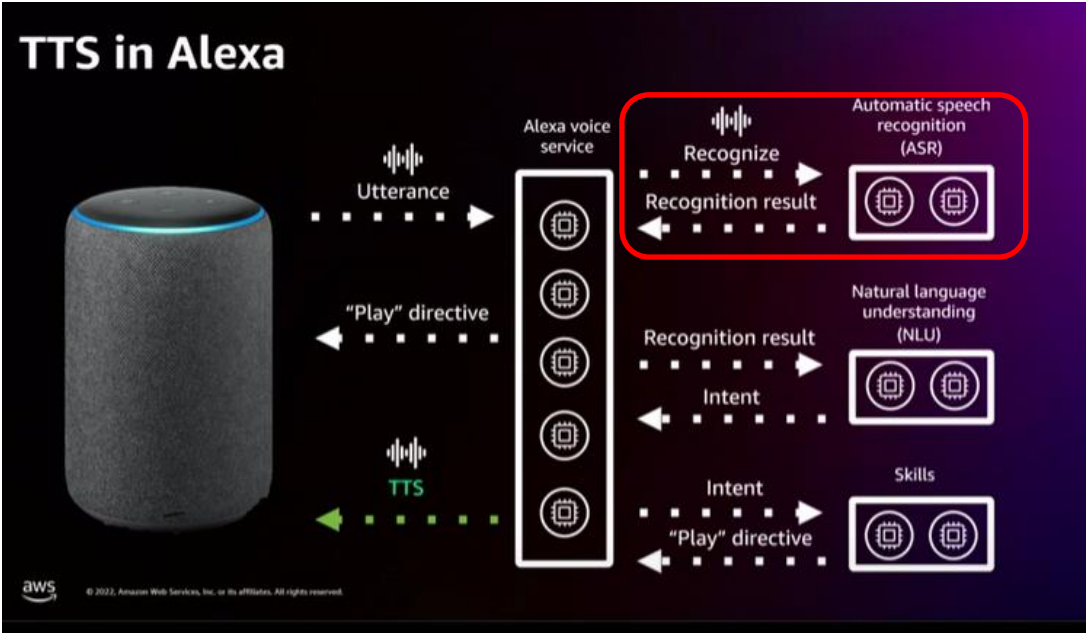
Exhibit J: Exemplary Infringement Claim Chart for U.S. Patent 9,502,025 Claim 8

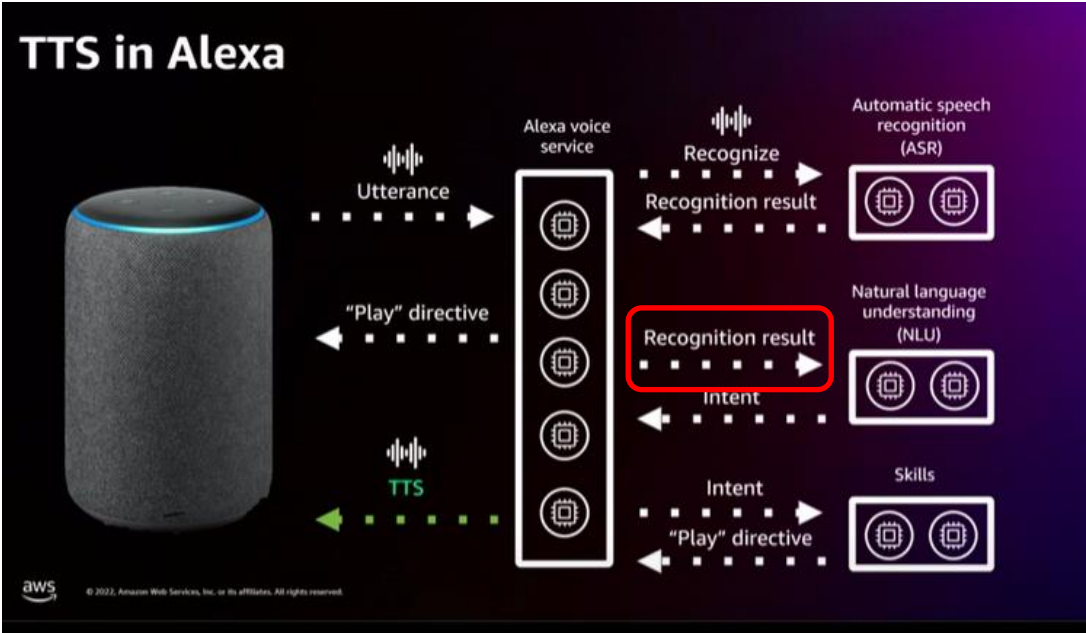
Claim	Exemplary Disclosures
<p>[8.pre] A method for providing a natural language content dedication service, the method being implemented in a computer system having one or more processors programmed with computer program instructions that, when executed by the one or more processors, cause the computer system to perform the method, the method comprising:</p>	<p>Amazon's Alexa is programmed to perform and have performed methods for providing a natural language content dedication service, the method being implemented in a computer system having one or more processors programmed with computer program instructions that, when executed by the one or more processors, cause the computer system to perform the method.</p> <p>Amazon's Alexa includes, for example, Echo, Echo Dot, Echo Show, FireTV, Alexa apps, and Alexa cloud.</p>  <p>See https://www.amazon.com/b?&node=9818047011&ref=MARS_NAV_desktop_plswitcher</p> <p>Alexa includes one or more processors. See https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/</p> <p>Additionally, Echo devices have one or more processors. For example, the Texas Instruments DM3725CUS100 Digital Media Processor as depicted below with red annotations:</p>

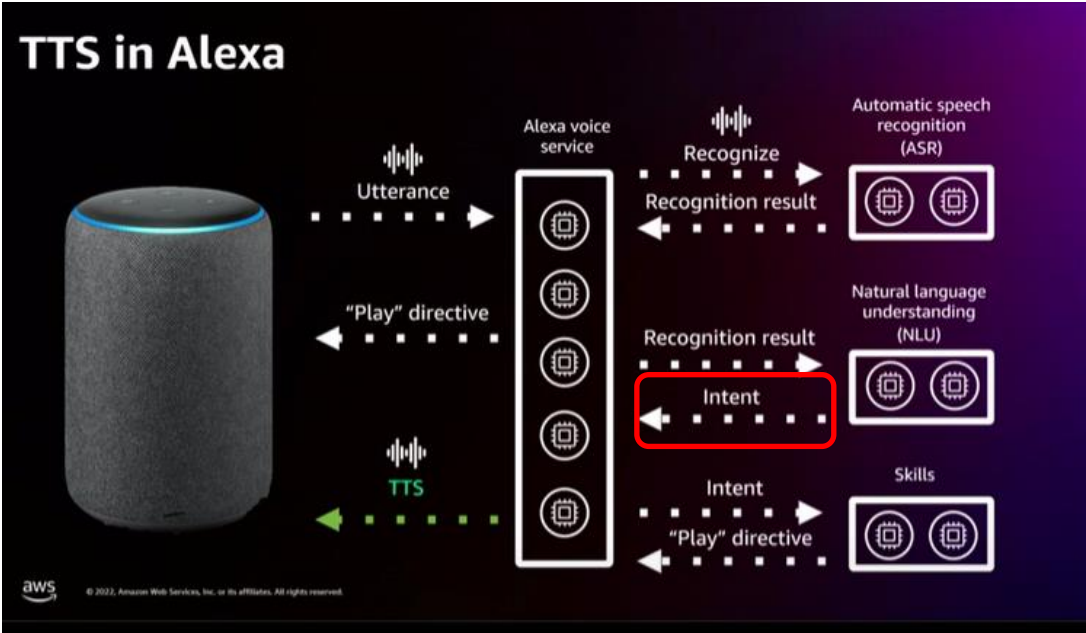
Claim	Exemplary Disclosures
	 <p data-bbox="688 724 1556 792">Sam Goldheart, Amazon Echo Teardown, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953</p>
[8.a] receiving, by one or more processors, a first utterance that includes a natural language utterance;	Alexa is programmed to perform and have performed methods for receiving a first utterance that includes a natural language utterance.

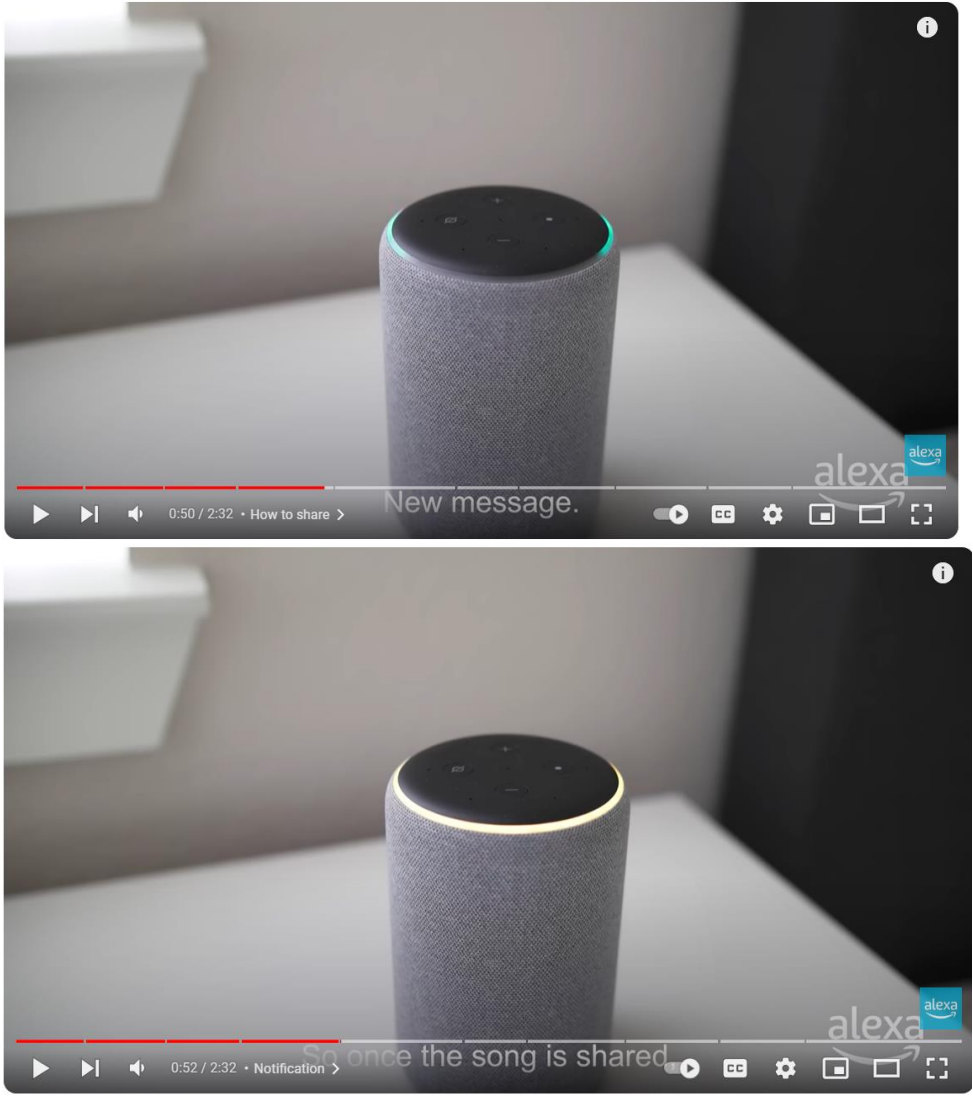
Claim	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p> <p>In the example shown below, Alexa receives an audio or transcribed version of the utterance: “Alexa share this song with [intended recipient].”</p>

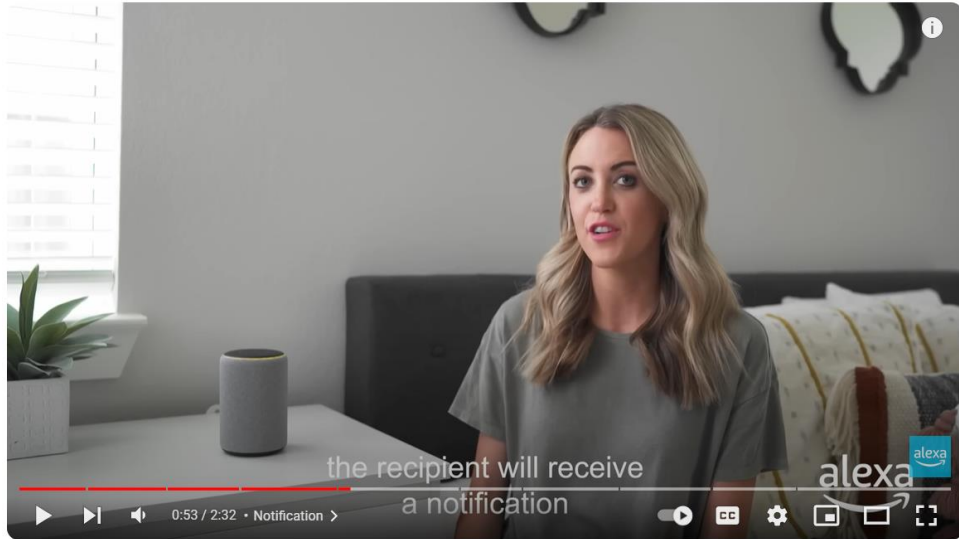
Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=rU5BBHOiWps</p> <p>See also Introducing Music Sharing, https://www.amazon.com/b?ie=UTF8&node=21568796011#:~:text=Everyone%20loves%20sharing%20music.,can%20send%20back%20a%20reaction “Now, customers can share songs with their Alexa contacts on their Echo devices. It’s as easy as saying, “Alexa share this song with Claire” while you’re listening to the song you want to share and Alexa will send the song to your chosen contact. The recipient can listen to the song using their Echo device and can send back a reaction.”</p>
<p>[8.b] determining, by the one or more processors, one or more words or phrases of the first utterance based on processing of the first utterance by a speech recognition engine;</p>	<p>Alexa is programmed to perform and have performed methods of determining, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance.</p>

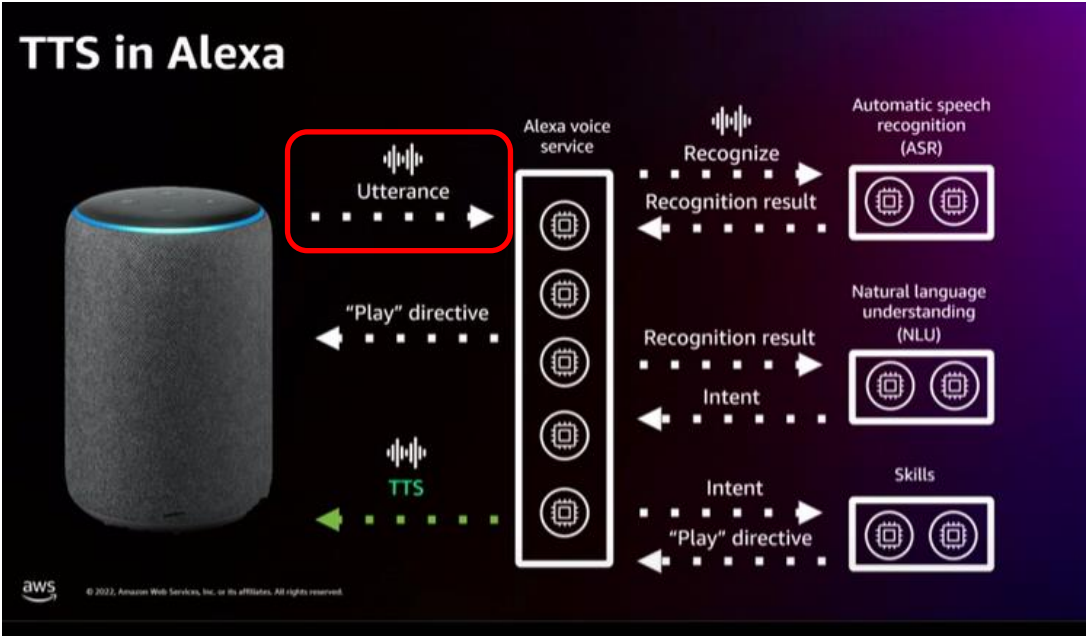
Claim	Exemplary Disclosures
	<p>TTS in Alexa</p>  <p>The diagram illustrates the TTS in Alexa process. It shows an Alexa voice service (represented by a vertical stack of four processor icons) interacting with an Alexa device (represented by a smart speaker). The process flow is as follows: 1. An 'Utterance' (represented by a waveform icon) is sent from the device to the Alexa voice service. 2. The 'Recognize' step is highlighted with a red box, showing the 'Utterance' being processed by 'Automatic speech recognition (ASR)' to produce a 'Recognition result'. 3. The 'Recognition result' is then sent to 'Natural language understanding (NLU)' to produce an 'Intent'. 4. The 'Intent' is then sent to 'Skills' to produce a '"Play" directive'. 5. The '"Play" directive' is sent back to the Alexa voice service, which then sends it to the device. 6. The device then uses 'TTS' (Text-to-Speech) to produce the final audio output, represented by a waveform icon.</p> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
[8.c] providing, by the one or more processors, the one or more words or phrases as an input to a conversational language processor;	Alexa is programmed to perform and have performed methods of providing the one or more words or phrases as an input to a conversational language processor.

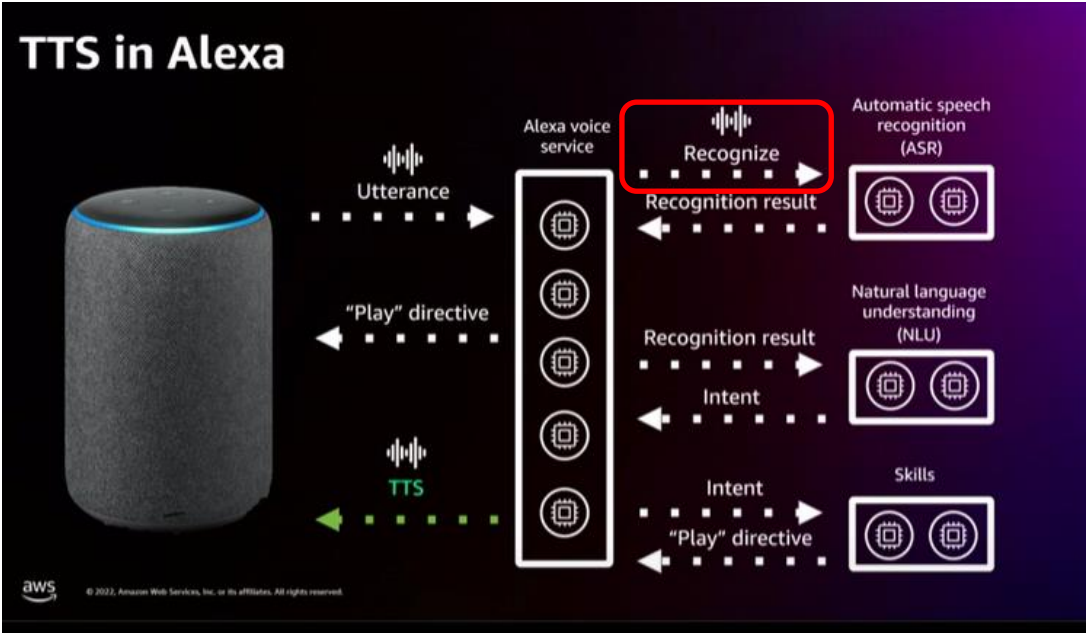
Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
<p>[8.d] interpreting, by the one or more processors, the first utterance, at the conversational language processor, based on the one or more words or phrases;</p> <p>[8.e] identifying, by the one or more processors, content to dedicate to a recipient based on the interpretation of the first utterance;</p>	<p>Alexa is programmed to perform and have performed methods of interpreting the first utterance, at the conversational language processor, based on the one or more words or phrases and identifying content to dedicate to a recipient based on the interpretation of the first utterance.</p>

Claim	Exemplary Disclosures
	 <p>The diagram, titled "TTS in Alexa", illustrates the process flow for a voice command. On the left is an image of an Alexa Echo device. A dashed arrow labeled "Utterance" points from the device to a central box labeled "Alexa voice service". This box contains four circular icons representing processors. From the "Alexa voice service", a dashed arrow labeled "Recognize" points to a box labeled "Automatic speech recognition (ASR)". A dashed arrow labeled "Recognition result" points back from ASR to the "Alexa voice service". Below this, another dashed arrow labeled "Recognize" points to a box labeled "Natural language understanding (NLU)". A dashed arrow labeled "Recognition result" points back from NLU to the "Alexa voice service". A red box highlights the "Intent" output from the NLU box, with a dashed arrow pointing to the "Alexa voice service". From the "Alexa voice service", a dashed arrow labeled "Intent" points to a box labeled "Skills". A dashed arrow labeled "Play" directive points back from the "Skills" box to the "Alexa voice service". Finally, a dashed arrow labeled "TTS" points from the "Alexa voice service" back to the Alexa device. The AWS logo and copyright notice are at the bottom left of the diagram.</p> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
<p>[8.f] initiating, by the one or more processors, a dedication to the recipient based on the identified content;</p>	<p>Alexa is programmed to perform and have performed methods of initiating, based on the identified content, a dedication to the recipient.</p> <p>Alexa sends a message about song to recipient.</p> <p>“Now, customers can share songs with their Alexa contacts on their Echo devices. It’s as easy as saying, “Alexa share this song with Claire” while you’re listening to the song you want to share and Alexa will send the song to your chosen contact. The recipient can listen to the song using their Echo device and can send back a reaction.”</p> <p>https://www.amazon.com/b?ie=UTF8&node=21568796011#:~:text=Everyone%20loves%20sharing%20music.,can%20send%20back%20a%20reaction.</p>

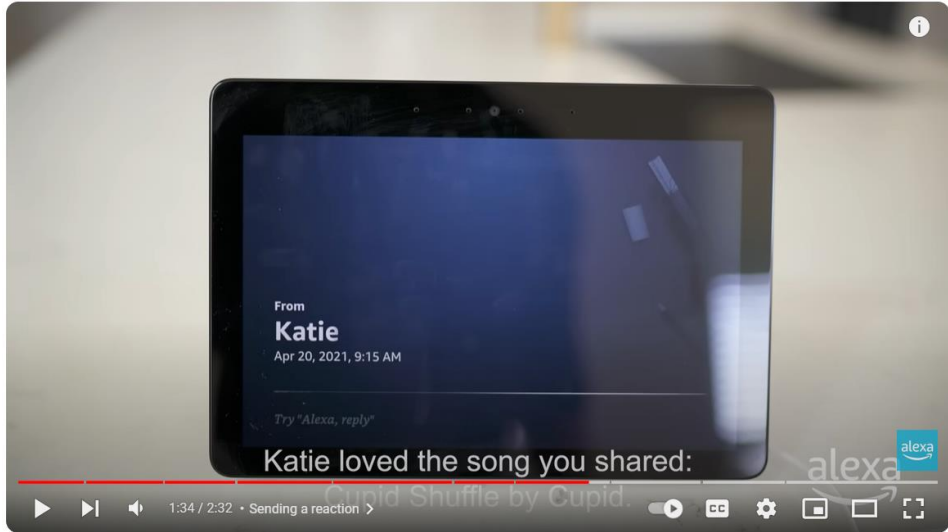
Claim	Exemplary Disclosures
	 <p>The image displays two sequential frames from a video. Both frames show a grey Amazon Echo speaker with a fabric mesh cover and a black top ring. In the top frame, the light ring is illuminated with a green glow. The bottom frame shows the same speaker with the light ring illuminated in yellow. Both frames include a video player interface at the bottom with a progress bar, play/pause button, volume icon, and other controls. The text 'New message.' is visible in the top frame, and 'So once the song is shared.' is visible in the bottom frame. The Alexa logo is also present in the bottom right corner of each frame.</p>

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=rU5BBHOiWps</p>
[8.g] receiving, by the one or more processors, a second utterance to be associated with the dedication;	Alexa is programmed to perform and have performed methods of receiving a second utterance to be associated with the dedication.

Claim	Exemplary Disclosures
	 <p>The diagram, titled "TTS in Alexa", illustrates the process of a voice interaction with an Alexa device. On the left, an Alexa Echo device is shown. A red box highlights the "Utterance" (represented by a waveform icon) being sent from the device to the "Alexa voice service" (a vertical stack of four processor icons). From the Alexa voice service, the process flows to "Automatic speech recognition (ASR)" and "Natural language understanding (NLU)", each represented by two processor icons. The flow is as follows: Utterance → Recognize → Recognition result → Intent → "Play" directive → Skills → "Play" directive → TTS → Alexa device. The "TTS" (Text-to-Speech) step is highlighted with a green arrow pointing back to the device. The AWS logo and copyright notice "© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved." are at the bottom left of the diagram.</p> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
<p>[8.h] determining, by the one or more processors, one or more words or phrases of the second utterance based on a processing of the second utterance by the speech recognition engine;</p>	<p>Alexa is programmed to perform or have performed methods of determining one or more words or phrases of the second utterance based on a processing of the second utterance by the speech recognition engine.</p>

Claim	Exemplary Disclosures
	 <p>The diagram, titled "TTS in Alexa", illustrates the process of a voice interaction with an Alexa device. On the left, an Alexa Echo device is shown. The process flow is as follows: <ul style="list-style-type: none"> An Utterance (represented by a soundwave icon) is sent from the device to the Alexa voice service (a vertical stack of four processor icons). The Alexa voice service sends a Recognize request (indicated by a red box around the soundwave icon) to Automatic speech recognition (ASR) (two processor icons). ASR returns a Recognition result to the Alexa voice service. The Alexa voice service sends another Recognize request to Natural language understanding (NLU) (two processor icons). NLU returns a Recognition result to the Alexa voice service. The Alexa voice service sends an Intent to Skills (two processor icons). Skills return a "Play" directive to the Alexa voice service. The Alexa voice service sends the "Play" directive back to the device. The device then performs TTS (Text-to-Speech, indicated by a green soundwave icon) to produce the final response. The AWS logo and copyright notice "© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved." are visible at the bottom left of the diagram. </p> <p>See https://www.youtube.com/watch?v=gRAFP--sZ_Y at 10:07.</p>
<p>[8.i] providing, by the one or more processors, the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and</p>	<p>Alexa is programmed to perform or have performed methods for providing, by the one or more processors, the one or more words or phrases of the second utterance as textual annotations within metadata of the content.</p> <p>“The Review Voice History section of Alexa Privacy Settings shows your voice interactions with Alexa. You can filter by date, device, and profile, and choose an entry to see what Alexa heard you say and how Alexa responded. You can even listen to the voice recording of each request. All of your voice recordings are saved by default, but you can change how long your voice recordings are saved or choose to delete them at any time.”</p> <p>https://www.amazon.com/b?ie=UTF8&node=23608614011</p>

Claim	Exemplary Disclosures
<p>[8.j] sending, by the one or more processors, information to enable the recipient to access the content and the second utterance.</p>	<p>Alexa is programmed to perform or have performed methods for sending, by the one or more processors, information to enable the recipient to access the content and the second utterance.</p>  <p>The image displays two sequential video frames of an Amazon Echo (Alexa) device. The device is a grey, cylindrical smart speaker with a glowing blue light ring at the top. It is placed on a white surface. The top frame shows a video player interface with the text 'Do you want to let Andrew know you like the song?' overlaid on the device. The bottom frame shows the same device with the text 'Yes!' overlaid. Both frames have an 'alexia' logo in the bottom right corner. The video player interface includes a progress bar, play/pause button, volume icon, and other standard controls.</p>

Claim	Exemplary Disclosures
	 <p>See https://www.youtube.com/watch?v=rU5BBHOiWps</p>