IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

VB Assets, LLC, Plaintiff, v. Amazon.com, Inc.; Amazon.com LLC; Amazon Web Services, Inc.; A2Z 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 Digital Services LLC, Defendants.	**C.A. No.** **JURY TRIAL DEMANDED**

COMPLAINT FOR PATENT INFRINGEMENT

VB Assets, LLC ("Plaintiff" or "VoiceBox") hereby alleges as follows for its complaint against Defendants Amazon.com, Inc.; Amazon.com LLC; Amazon Web Services, Inc.; A2Z 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 Digital Services LLC (collectively, "Defendants" or "Amazon"):

NATURE OF THE ACTION

1. VoiceBox, through its predecessor companies VoiceBox Technologies Corporation and VoiceBox Technologies, Inc. (collectively "VoiceBox Technologies"), pioneered voice-based search and commerce technology. It invented what Amazon itself has described as "Echo-like" products long before Amazon.

2. In recognition of its many innovations, the U.S. Patent & Trademark Office awarded and issued the VoiceBox Patents.[1] The innovations in these patents were fundamental to the development of voice commerce technology.

3. VoiceBox Technologies' opportunities to promote and build a business based on these patents were crushed when Amazon introduced the infringing Echo and Alexa Products[2] and used its enormous size and clout to poach dozens of VoiceBox Technologies' engineers and scientists.

4. VoiceBox has brought this case to hold Amazon accountable for its infringement of VoiceBox's patent rights.

---

[1] "VoiceBox Patents" collectively refers to U.S. Patent Nos. 8,073,681 ("the '681 patent"); 9,015,049 ("the '049 patent"); 9,626,703 ("the '703 patent"); 7,818,176 ("the '176 patent"); 8,886,536 ("the '536 patent"); and 9,269,097 ("the '097 patent").

[2] "Alexa Products" collectively refers to Amazon's Alexa virtual assistant and offerings that include Alexa, including the Echo product line (such as Echo 1st Gen., Echo 2nd Gen., Echo Dot 1st Gen., Echo Dot 2nd Gen., Echo Dot 3rd Gen., Echo Dot Kids Edition, Echo Show 1st Gen., Echo Show 2nd Gen., Echo Show 5, Echo Spot, Echo Plus 1st Gen., Echo Plus 2nd Gen., Echo Auto, and Echo Look), Amazon's Alexa apps, Music apps, and Shopping apps on a smartphone or other mobile device, Amazon's Alexa cloud, Alexa Voice Services, and Amazon.com website, and any other device, app, or instrumentality that includes, provides access to, or works with Alexa (such as Amazon Tap, Amazon Dash Wand, Echo Wall Clock, AmazonBasics Microwave, Amazon SmartPlug, Amazon Fire TV Sticks, Amazon Fire TVs, Amazon Fire TV Cubes, and Amazon Fire and Fire HD tablets) as well as software, hardware, and cloud infrastructure associated with any of the foregoing.

THE PARTIES

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

6. On information and belief, Amazon.com, Inc. is a corporation incorporated in Delaware and has a principal place of business at 410 Terry Avenue North, Seattle, WA, 98109. On information and belief, Amazon.com, Inc. is the ultimate parent company of the other companies that make up Amazon, and is responsible for making, using, selling, offering for sale and/or importing Alexa Products.

7. On information and belief, Amazon.com LLC is a limited liability company organized under the laws of Delaware and has a principal place of business at 410 Terry Avenue North, Seattle, WA, 98109. On information and belief, Amazon.com LLC includes as members various operating companies, which make, use, sell, offer for sale, and/or import Alexa Products.

8. On information and belief, Amazon Web Services, Inc. is a corporation incorporated in Delaware and has a principal place of business at 410 Terry Avenue North, Seattle, WA, 98109. On information and belief, Amazon Web Services, Inc. provides Alexa cloud computing platforms, that are Alexa Products or are for use with one or more Alexa Products.

9. On information and belief, A2Z Development Center, Inc. d/b/a Lab126 is a corporation incorporated in Delaware and has its principal place of business at 1120 Enterprise Way Sunnyvale, CA 94089. On information and belief, A2Z Development Center, Inc. d/b/a Lab126 performed research and development, including for one or more Alexa Products.

10. On information and belief, Rawles LLC is a limited liability company formed under the laws of Delaware and has its principal place of business at 103 Foulk Road, Suite 100,

Wilmington, DE 19803. On information and belief, Rawles LLC performed research and development, including for one or more Alexa Products.

11. On information and belief, AMZN Mobile LLC is a limited liability company organized under the laws of Delaware and has a principal place of business near Seattle, WA. On information and belief, AMZN Mobile LLC provides mobile apps that are Alexa Products or are for use with Alexa Products.

12. On information and belief, AMZN Mobile 2 LLC is a limited liability company organized under the laws of Delaware and has a principal place of business near Seattle, WA. On information and belief, AMZN Mobile 2 LLC provides mobile apps that are Alexa Products or are for use with Alexa Products.

13. On information and belief, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc. is a corporation incorporated in Delaware and has a principal place of business at 410 Terry Ave. N. Seattle, WA 98109. On information and belief, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc. participated in the sale or offer for sale of one or more Alexa Products.

14. On information and belief, Amazon Digital 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 Digital 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

15. 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.

16. Personal jurisdiction over each Defendant 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. Each Defendant 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.

17. 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), each Defendant 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.

FACTUAL BACKGROUND

A. VoiceBox Technologies Invents Groundbreaking Voice Technology

18. In 2001, three brothers, Mike, Rich, and Bob Kennewick founded VoiceBox Technologies 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 Technologies could become the first company to enable people to naturally and effectively interact with computer speech systems.

19. From its inception, VoiceBox Technologies engaged in intense research efforts to develop its NLU technology. As part of these efforts, VoiceBox Technologies achieved a

significant milestone when it developed an early prototype called "Cybermind." As demonstrated on the King5 news, Cybermind was a voice-controlled speaker that could provide weather, recipes, sports scores, calendar updates, or play a song.[3]

Figure 1: Cybermind Prototype

20. VoiceBox Technologies' groundbreaking work did not go unrecognized. After learning about VoiceBox Technologies' technology, Toyota hired it to build a sophisticated NLU speech interface for its Lexus automobiles. VoiceBox Technologies built the voice and NLU capability for Toyota's award-winning Entune multimedia system.

21. As part of the development effort of an NLU interface for Lexus, VoiceBox Technologies demonstrated a personal assistant called "Alexus" that showcased the power of its Conversational Voice technology. On information and belief, the VoiceBox Technologies "Alexus" concept was introduced to the public more than six months before Amazon announced "Alexa."

[3] https://www.youtube.com/watch?v=DDcRyPnvWhw

Figure 2: "Alexus" Demonstration

22. Throughout its research and development efforts, VoiceBox Technologies realized that its technology could be deployed in a wide range of applications from connected home to mobile personal assistants.

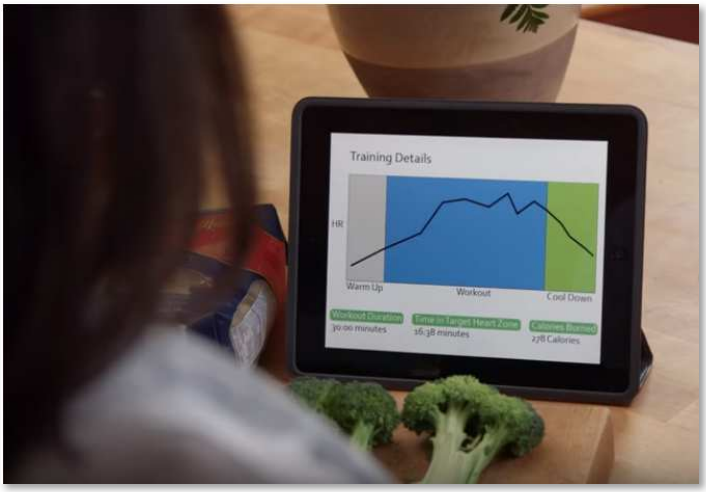


Figure 3: Connected Home

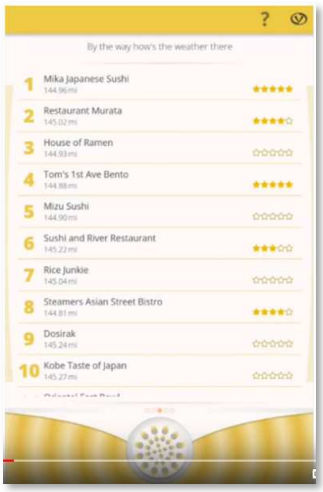


Figure 4: Mobile

23. By January 2012, VoiceBox Technologies was a leader in NLU and conversational voice technology. Leading companies throughout the world, including Toyota, Lexus, TomTom, Pioneer, Chrysler, Dodge, and Magellan used VoiceBox Technologies' award-winning and patented contextual speech technology. VoiceBox Technologies had software

applications that ran on smart speakers, in-car systems, smartphones, smart TVs, computers, tablets, e-readers, and personal navigation devices.

24. In 2013, the Institute of Electrical and Electronics Engineers ("IEEE") ranked VoiceBox Technologies number 13 in patent power for the computer software industry. VoiceBox Technologies had become the leader in conversational Artificial Intelligence ("AI"), including Voice Recognition (VR), NLU, and AI services.

B. Amazon Takes VoiceBox's Technology

25. In 2011, VoiceBox Technologies contacted Amazon to explore a potential business relationship where VoiceBox Technologies 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 Technologies provided Amazon with a presentation that described its award-winning patented technology and explicitly referred to VoiceBox Technologies' "patented Contextual Speech Technology." Slides from that presentation are reproduced below:

Figure 5: 2011 Slides

26. Amazon was so impressed by the technology VoiceBox Technologies presented on October 7, 2011 that its representative emailed VoiceBox Technologies 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 Technologies'] personal digital assistant and underlying conversational voice technology." On information and belief, in addition to Mr. Booms, the following Amazon personnel attended the meeting: Nick Komorous (Director, Corporate Development), Ian Freed (VP, Amazon Devices), Greg Hart (VP, Digital), Al Lindsay (Director, Software Engineer), Frederic Deramat (Senior Principal Engineer), and John Thimsen (Software Developer). On information and belief, another Amazon engineer from Cupertino attended by teleconference.

27. Two days after the meeting at Amazon, on Friday October 21, 2011, Amazon's Mr. Komorous emailed VoiceBox Technologies and asked to visit the company's office for a "deeper dive." Mr. Komorous requested that this meeting occur the very next week. VoiceBox Technologies agreed to host Amazon's personnel for a meeting at VoiceBox Technologies' office on October 26, 2011. On information and belief, from Amazon, Marcello Typrin (Product Management / Business Development), Frederic Deramat (Senior Principal Engineer), John Thimsen (Software Developer), and Sean Fitz (Software Developer) joined Mr. Komorous at the meeting.

28. 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 indicated that Amazon's culture was "engineering heavy" and asked that VoiceBox Technologies "have engineering and speech representation at the meeting."

29. The October 26, 2011 meeting at VoiceBox Technologies' office ran from 10 am until 12 noon with some Amazon engineers staying even later until around 2:30 pm. The meeting included a tour with additional demos, a review of the technical architecture for VoiceBox Technologies' server software, and plans for next generation products.

30. During the meeting, VoiceBox Technologies presented a deck of 42 slides to the visiting Amazon personnel. The slides provided even more detail about VoiceBox Technologies' patented technology and informed Amazon that VoiceBox Technologies had 12 patents at the time with an additional 14 pending applications. The slide deck included the following slide regarding the '176 patent, which is asserted in this lawsuit.

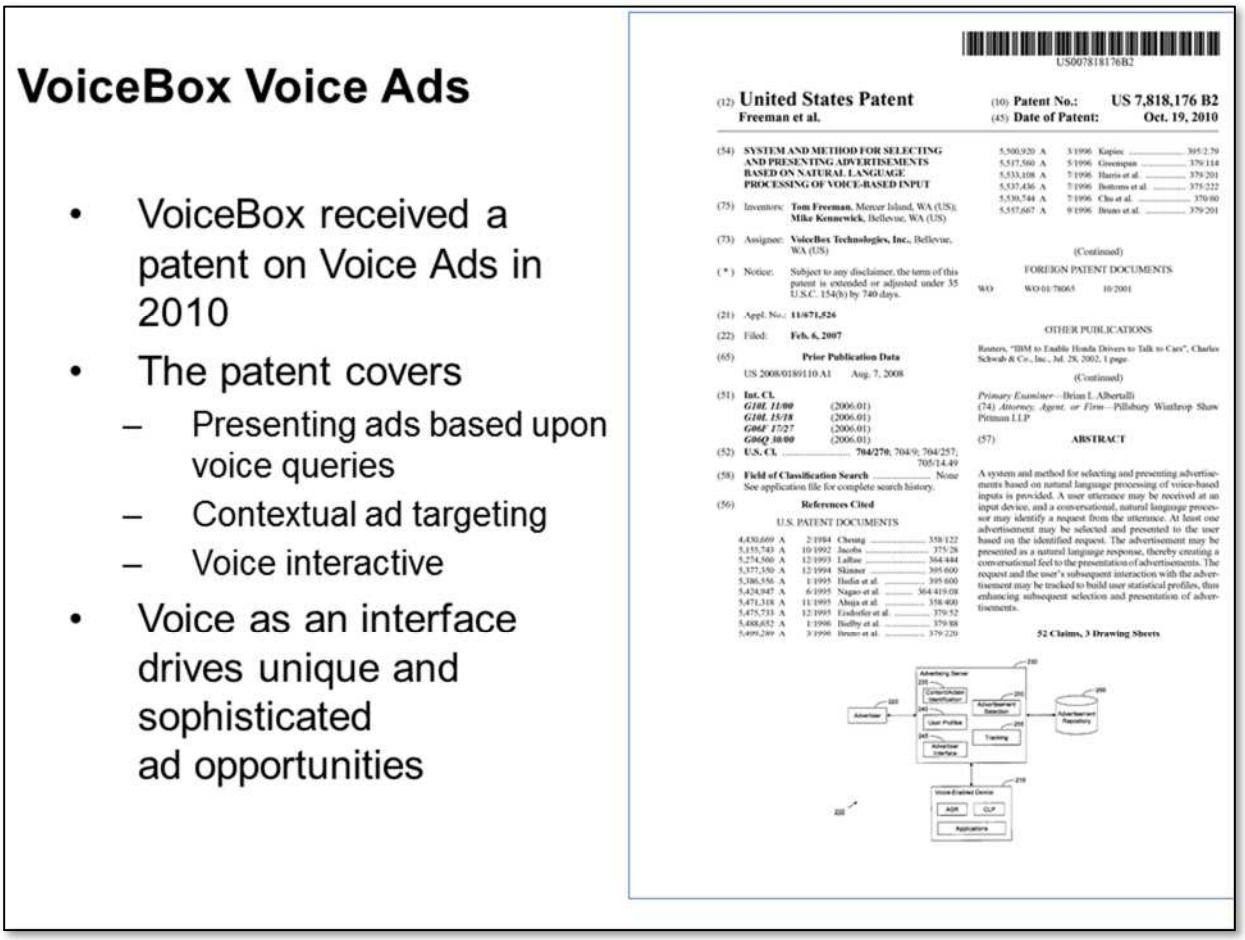


Figure 6: Slide from 2011 Presentation

31. The slide deck VoiceBox Technologies presented to Amazon at the October 26, 2011 meeting also proposed a business arrangement where VoiceBox Technologies would provide "Voice Services" to Amazon. The VoiceBox Technologies Voice Services from the presentation are shown below:

Figure 7: Slide from 2011 Presentation

32. On information and belief, some of the Amazon personnel involved in the 2011 meetings became technical leaders for Amazon's Alexa Products while others became high-level executives with close working relationships with the senior leadership of Amazon. On information and belief, one of Mr. Freed's past positions was Technical Advisor to CEO, a position that has been described as CEO Jeff Bezos's "shadow" advisor. On information and

belief, Mr. Freed then went on to become 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 CEO position around 2011. On information and belief, around 2011, Al Lindsey was promoted to Vice President managing the Alexa Engine Software team. On information and belief, around 2011, Mr. Deramat was promoted to the position Vice President & Distinguished Engineer for Amazon Alexa. On information and belief, around 2011, Mr. Thimsen was promoted to the position 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."

33. A couple days after the last meeting in 2011, VoiceBox Technologies 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 [Amazon's] future." Amazon did not provide the results of these discussions and did not pursue a business relationship with VoiceBox Technologies. Instead, on information and belief, Amazon decided to build its Alexa Products—without telling VoiceBox Technologies or asking permission to use VoiceBox Technologies' patented technology.

34. In 2014, Amazon announced the launch of Alexa, a virtual assistant, along with the first-generation Echo product, a smart speaker. Amazon's Alexa and first-generation Echo product were strikingly similar to the patented technology that VoiceBox Technologies showed Amazon in 2011.

35. In 2016, Amazon abruptly hired Philippe Di Cristo, who was VoiceBox Technologies' Chief Scientist. While at VoiceBox Technologies, Mr. Di Cristo gained knowledge

of the company's voice technology and had full access to VoiceBox Technologies' intellectual property. As Mr. Di Cristo explains on his LinkedIn Page, he had worked on an "Amazon Echo-like system" while at VoiceBox Technologies. Exhibit A at 3. On information and belief, Mr. Di Cristo has helped design and implement VoiceBox Technologies' patented technology into Amazon's Echo and other Alexa Products.

36. On information and belief, Mr. Di Cristo played an active role in soliciting additional VoiceBox employees to join Amazon. For example, shortly after Mr. Di Cristo joined Amazon from VoiceBox Technologies, Amazon ramped-up its efforts to recruit VoiceBox Technologies employees.

37. On January 10, 2017, Amazon hosted 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." The director of Amazon's Voice & Advanced NUI Shopping group— on information and belief, the group Mr. Di Cristo joined—sent an invite for the event to a large number of VoiceBox Technologies employees and indicated 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." For this event, Amazon rented out Seastar, the premier seafood restaurant near VoiceBox Technologies' office.

38. On information and belief, Mr. Di Cristo was originally scheduled to be a speaker at the event targeting VoiceBox Technologies employees—but abruptly withdrew. The event emphasized Amazon's need for VoiceBox Technologies' NLU technology. On information and belief, around this time, Amazon faced numerous NLU challenges and had only completed a

small portion of work toward Amazon's goal to build conversational dialog interaction into Alexa Products.

39. On January 17, 2017, after Amazon's rampant poaching of VoiceBox Technologies employees came to light, Mike Kennewick, CEO of VoiceBox Technologies, sent a letter to Jeff Bezos, CEO of Amazon, to propose a business solution. A true and correct copy of that letter is attached as Exhibit B. The letter explains that, at the time, VoiceBox Technologies 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."

40. On information and belief, Mr. Booms, Amazon's Vice President of Worldwide Corporate Development, responded to VoiceBox Technologies' January 17, 2017 letter on behalf of Mr. Bezos. In Amazon's response, Mr. Booms requested a meeting at VoiceBox Technologies' offices for purposes of "go[ing] fairly deep on the technology, data, customer relationships."

41. On February 2, 2017, the parties met at VoiceBox Technologies' office. Amazon came with a team of technologists from its Alexa program, including on information and belief, Manoj Sindhwani (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 Technologies provided a detailed technical presentation, which included information about patents and pending applications then owned by VoiceBox Technologies. Slides in the presentation listed all VoiceBox Technologies patents and published applications at the time, including the '681 patent, the '049 patent, the '176 patent, the '536 patent, and the '097 patent. The February 2, 2017 slides also reproduced a highlighted claim from the '681 patent and a highlighted claim from the '176 patent.

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

43. In March of 2017, VoiceBox Technologies hosted another meeting with Amazon personnel at VoiceBox Technologies' office. The Amazon attendees included senior executives and managers responsible for Alexa. VoiceBox Technologies informed Amazon by email after the meeting that VoiceBox Technologies had "[p]atents that could be useful as the market goes mainstream." Shortly, thereafter, Mr. Komorous from Amazon emailed VoiceBox Technologies, writing that Amazon had been "[poring] through the material" VoiceBox Technologies provided and that Amazon had created yet another, even more detailed, list of requests.

44. In April of 2017, VoiceBox Technologies shared a written summary of VoiceBox Technologies' patent portfolio with Amazon. The written summary explained that VoiceBox Technologies' patents covered "core & fundamental areas." The summary showcased the '176 patent and the '703 patent, and identified the '536 patent and '097 patent by number. The summary further explained that "[v]oice advertising & conversational e-commerce are essential to the core strategies and future revenue streams of many of the world's leading technology companies, including: . . . Amazon . . . ."

OVERVIEW OF VOICEBOX'S PATENTS-IN-SUIT

45. United States Patent Number 8,073,681 ("the '681 patent"), entitled "System and Method for a Cooperative Conversational Voice User Interface," was duly and legally issued on December 6, 2011, and names Larry Baldwin, Tom Freeman, Michael Tjalve, Blane Ebersold, and Chris Weider as the inventors. Attached as Exhibit C is a true and correct copy of the '681 patent.

46. The ’681 patent claims, among other things, a system for providing a cooperative conversational voice user interface, comprising: a voice input device configured to receive an utterance during a current conversation with a user; and a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to: accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation; accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user; identify a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge; infer additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context; establish an intended meaning for the utterance within the identified identify a context based on the additional information inferred about the utterance; and generate a response to the utterance based on the intended meaning established within the identified context.

47. VoiceBox is the assignee of the entire right, title, and interest in the ’681 patent.

48. United States Patent Number 9,015,049 (“the ’049 patent”), entitled “System and Method for a Cooperative Conversational Voice User Interface,” was duly and legally issued on April 21, 2015, and names Larry Baldwin, Tom Freeman, Michael Tjalve, Blane Ebersold, and Chris Weider as the inventors. Attached as Exhibit E is a true and correct copy of the ’049 patent.

49. The ’049 patent claims, among other things, a system for facilitating conversation-based responses, the system comprising: one or more physical processors programmed with one or more computer program instructions such that, when executed, the one

or more computer program instructions cause the one or more physical processors to: receive a natural language utterance during a conversation between a user and the system; identify a first model that includes short-term knowledge about the conversation, wherein the short-term knowledge is based on one or more prior natural language utterances received during the conversation; identify, based on the short-term knowledge, context information for the natural language utterance; determine, based on the context information, an interpretation of the natural language utterance; and generate, based on the interpretation of the natural language utterance, a response to the natural language utterance.

50. VoiceBox is the assignee of the entire right, title, and interest in the ’049 patent.

51. United States Patent Number 9,626,703 (“the ’703 patent”), entitled “Voice Commerce,” was duly and legally issued on April 18, 2017, and names Michael R. Kennewick, Sr. as the inventor. Attached as Exhibit G is a true and correct copy of the ’703 patent.

52. The ’703 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 user input comprising a natural language utterance; provide, without further user input after the receipt of the user input, the natural language utterance as an input to a speech recognition engine; obtain, without further user input after the receipt of the user input, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine; identify, without further user input after the receipt of the user input, a context based at least on the one or more words or phrases; determine, without further user input identifying a product or service after the receipt of the user input, a first product or service to be purchased on behalf of a user based at least on the determined context; obtain, without further

user input identifying payment information after the receipt of the user input, first payment information with which to pay for the product or service; obtain, without further user input identifying shipping information after the receipt of the user input, first shipping information with which to deliver the product or service, wherein the first shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased; and complete, without further user input identifying a product or service, payment information, or shipping information after the receipt of the user input, a purchase transaction for the first product or service based on the first payment information and the first shipping information.

53. VoiceBox is the assignee of the entire right, title, and interest in the ’703 patent.

54. United States Patent Number 7,818,176 (“the ’176 patent”), entitled “System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-based Input,” was duly and legally issued on October 19, 2010, and names Tom Freeman and Mike Kennewick as the inventors. Attached as Exhibit I is a true and correct copy of the ’176 patent.

55. The ’176 patent claims, among other things, a system for selecting and presenting advertisements in response to processing natural language utterances, comprising: an input device that receives a natural language utterance containing at least one request at an input device; a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to: map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and generate a

preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases; a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; select an advertisement in the context established for the natural language utterance; and present the selected advertisement via an output device.

56. VoiceBox is the assignee of the entire right, title, and interest in the '176 patent.

57. United States Patent Number 8,886,536 ("the '536 patent"), entitled "System and Method for Delivering Targeted Advertisements and Tracking Advertisement Interactions in Voice Recognition Contexts," was duly and legally issued on November 11, 2014, and names Tom Freeman and Mike Kenn[e]wick as the inventors. Attached as Exhibit K is a true and correct copy of the '536 patent.

58. The '536 patent claims, among other things, a computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance; providing, by the one or more physical processors, a response to the first natural language utterance; receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance; identifying, by the one or more physical processors, requests associated with the second natural

language utterance, wherein the requests include a first request to be processed by a first device associated with a user and a second request to be processed by a second device associated with the user; determining, by the one or more physical processors, promotional content that relates to one or more of the first request or the second request; and presenting, by the one or more physical processors, the promotional content to the user.

59. VoiceBox is the assignee of the entire right, title, and interest in the ’536 patent.

60. United States Patent Number 9,269,097 (“the ’097 patent”), entitled “System and Method for Delivering Targeted Advertisements and/or Providing Natural Language Processing Based on Advertisements,” was duly and legally issued on February 23, 2016, and names Tom Freeman and Mike Kennewick as the inventors. Attached as Exhibit M is a true and correct copy of the ’097 patent.

61. The ’097 patent claims, among other things, a method for providing natural language processing based on advertisements, the method being implemented on a computer system having one or more physical processors executing computer program instructions which, when executed, perform the method, the method comprising: providing, by the computer system, an advertisement associated with a product or service for presentation to a user; receiving, at the computer system, a natural language utterance of the user; and interpreting, by the computer system, the natural language utterance based on the advertisement and, responsive to the existence of a pronoun in the natural language utterance, determining whether the pronoun refers to one or more of the product or service or a provider of the product or service.

62. VoiceBox is the assignee of the entire right, title, and interest in the ’097 patent.

COUNT ONE (INFRINGEMENT OF '681 PATENT)

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

64. Amazon has and continues to infringe the '681 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 '681 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 D.

65. Amazon has been and is inducing infringement of the '681 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 '681 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 '681 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '681 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '681 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, the '681 patent or its published

application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent Nos. 9,424,840; 9,852,729; 9,922,639; 10,026,394; and 10,102,845. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '681 patent since no later than March 4, 2015, the date on which the '681 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

66. Amazon has been and is continuing to contributorily infringe the '681 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '681 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 '681 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '681 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '681 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, the '681 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent Nos. 9,424,840; 9,852,729; 9,922,639; 10,026,394; and 10,102,845. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '681 patent since no later than March 4, 2015, the date on which the '681 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon. 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.

67. On information and belief, Amazon has known of the existence of the '681 patent, and its acts of infringement have been willful and in disregard for the '681 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 '681 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 '681 patent. Additionally, VoiceBox Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '681 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '681 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '681 patent. Additionally, the '681 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent Nos. 9,424,840; 9,852,729; 9,922,639;

10,026,394; and 10,102,845. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '681 patent since no later than March 4, 2015, the date on which the '681 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '681 patent.

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

69. Amazon's acts of infringement of the '681 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

70. 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 '049 PATENT)

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

72. Amazon has and continues to infringe the '049 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 '049 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.

73. Amazon has been and is inducing infringement of the '049 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 '049 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 '049 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '049 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '049 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, the '681 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent Nos. 9,424,840; 9,852,729; 9,922,639; 10,026,394; and 10,102,845. The '681 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on March 4, 2015. By March 4, 2015, the '049 patent's application had published claiming priority to the '681 patent. On April 21, 2015, the '049 patent issued claiming priority to the '681 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '049 patent since no later than April 21, 2015. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

74. Amazon has been and is continuing to contributorily infringe the '049 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially

adapted for practicing the invention of the '049 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 '049 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '049 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '049 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, the '681 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent Nos. 9,424,840; 9,852,729; 9,922,639; 10,026,394; and 10,102,845. The '681 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on March 4, 2015. By March 4, 2015, the '049 patent's application had published claiming priority to the '681 patent. On April 21, 2015, the '049 patent issued claiming priority to the '681 patent. On information and belief,

Amazon has known of, or been willfully blind to, the existence of the ’049 patent since no later than April 21, 2015. 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.

75. On information and belief, Amazon has known of the existence of the ’049 patent, and its acts of infringement have been willful and in disregard for the ’049 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 ’049 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 ’049 patent. Additionally, VoiceBox Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies’ office of the ’049 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’049 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the ’049 patent. Additionally, the ’681 patent or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent Nos. 9,424,840; 9,852,729; 9,922,639; 10,026,394; and 10,102,845. The ’681 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on March 4, 2015. By March 4, 2015, the ’049 patent’s application had published claiming priority to the ’681 patent. On April 21, 2015, the ’049 patent issued claiming priority to the ’681 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’049 patent since no later than April 21, 2015. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the ’049 patent.

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

77. Amazon's acts of infringement of the '049 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

78. 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 '703 PATENT)

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

80. Amazon has and continues to infringe the '703 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 '703 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.

81. Amazon has been and is inducing infringement of the '703 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 '703 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 ’703 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 April of 2017, VoiceBox Technologies sent a written summary of VoiceBox Technologies’ patent portfolio to Amazon that informed Amazon of the ’703 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’703 patent since no later than April of 2017. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

82. Amazon has been and is continuing to contributorily infringe the ’703 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the ’703 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 ’703 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 April of 2017, VoiceBox Technologies sent a written summary of VoiceBox Technologies’ patent portfolio to Amazon that informed Amazon of the ’703 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’703 patent since no later

than April of 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.

83. On information and belief, Amazon has known of the existence of the '703 patent, and its acts of infringement have been willful and in disregard for the '703 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 '703 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 '703 patent. Additionally, in April of 2017, VoiceBox Technologies sent a written summary of VoiceBox Technologies' patent portfolio to Amazon that informed Amazon of the '703 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '703 patent since no later than April of 2017. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '703 patent.

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

85. Amazon's acts of infringement of the '703 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

86. 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 '176 PATENT)

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

88. Amazon has and continues to infringe the '176 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 '176 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.

89. Amazon has been and is inducing infringement of the '176 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 '176 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 '176 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 notified the Amazon employees that attended the October 26, 2011 meeting at VoiceBox Technologies' office of the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '176 patent since October 26, 2011. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, the '176 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. 10,152,973. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '176 patent since no later

than August 10, 2018, the date on which the '176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. By that date, U.S. Pat. No. 8,145,489 had issued claiming priority to the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '176 patent since no later than January 12, 2015. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

90. Amazon has been and is continuing to contributorily infringe the '176 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the '176 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 '176 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 notified the Amazon employees that attended the October 26, 2011 meeting at VoiceBox Technologies' office of the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '176 patent since October 26, 2011. 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, the '176 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. 10,152,973. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '176 patent since no later than August 10, 2018, the date on which the '176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon. 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, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. By that date, U.S. Pat. No. 8,145,489 had issued claiming priority to the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '176 patent since no later than January 12, 2015. 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.

91. On information and belief, Amazon has known of the existence of the ’176 patent, and its acts of infringement have been willful and in disregard for the ’176 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 ’176 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 ’176 patent. Additionally, VoiceBox notified the Amazon employees that attended the October 26, 2011 meeting at VoiceBox Technologies’ office of the ’176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’176 patent since no later than October 26, 2011. No later than that date, Amazon has known its Alexa Products infringe the ’176 patent. Additionally, the ’176 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. 10,152,973. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’176 patent since no later than August 10, 2018, the date on which the ’176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the ’176 patent. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. By that date, U.S. Pat. No. 8,145,489 had issued claiming priority to the ’176 patent. On information and belief, Amazon has known

of, or been willfully blind to, the existence of the '176 patent since no later than January 12, 2015. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '176 patent.

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

93. Amazon's acts of infringement of the '176 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

94. 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 '536 PATENT)

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

96. Amazon has and continues to infringe the '536 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 '536 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 L.

97. Amazon has been and is inducing infringement of the '536 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 '536 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 '536 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '536 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '536 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, the '176 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. 10,152,973. The '176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on August 10, 2018. As of that date, the '536 patent had issued claiming priority to the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '536 patent since no later than August 10, 2018. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. By that date, the '536 patent had issued claiming priority to U.S. Pat. No. 8,145,489. On information and belief, Amazon has known of, or been willfully blind to,

the existence of the ’536 patent since no later than January 12, 2015. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

98. Amazon has been and is continuing to contributorily infringe the ’536 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the ’536 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 ’536 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies’ office of the ’536 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’536 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, the ’176 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. 10,152,973. The ’176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on August 10, 2018. As of that date, the ’536 patent had issued claiming priority to the ’176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’536 patent since no later than August 10, 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. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. By that date, the ’536 patent had issued claiming priority to U.S. Pat. No. 8,145,489. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’536 patent since no later than January 12, 2015. 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.

99. On information and belief, Amazon has known of the existence of the ’536 patent, and its acts of infringement have been willful and in disregard for the ’536 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 ’536 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 ’536 patent. Additionally, VoiceBox Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies’ office of the

'536 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '536 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '536 patent. Additionally, the '176 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. 10,152,973. The '176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on August 10, 2018. As of that date, the '536 patent had issued claiming priority to the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '536 patent since no later than August 10, 2018. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '536 patent. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. By that date, the '536 patent had issued claiming priority to U.S. Pat. No. 8,145,489. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '536 patent since no later than January 12, 2015. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '536 patent.

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

101. Amazon's acts of infringement of the '536 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

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

COUNT SIX (INFRINGEMENT OF THE '097 PATENT)

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

104. Amazon has and continues to infringe the '097 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 '097 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 N.

105. Amazon has been and is inducing infringement of the '097 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 '097 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 '097 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '097 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the

’097 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, the ’176 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. 10,152,973. The ’176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on August 10, 2018. As of that date, the ’097 patent had issued claiming priority to the ’176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’097 patent since no later than August 10, 2018. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. On February 23, 2016, the ’097 patent issued claiming priority to U.S. Pat. No. 8,145,489. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’097 patent since no later than February 23, 2016. On information and belief, no later than that date, Amazon has known its actions would induce infringement by users of Alexa Products.

106. Amazon has been and is continuing to contributorily infringe the ’097 patent by selling or offering to sell Alexa Products, knowing them to be especially made or especially adapted for practicing the invention of the ’097 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 '097 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 Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '097 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '097 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, the '176 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. 10,152,973. The '176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on August 10, 2018. As of that date, the '097 patent had issued claiming priority to the '176 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '097 patent since no later than August 10, 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. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during

prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. On February 23, 2016, the '097 patent issued claiming priority to U.S. Pat. No. 8,145,489. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '097 patent since no later than February 23, 2016. 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.

107. On information and belief, Amazon has known of the existence of the '097 patent, and its acts of infringement have been willful and in disregard for the '097 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 '097 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 '097 patent. Additionally, VoiceBox Technologies informed the Amazon employees who attended the February 2, 2017 meeting at VoiceBox Technologies' office of the '097 patent. On information and belief, Amazon has known of, or been willfully blind to, the existence of the '097 patent since no later than February 2, 2017. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the '097 patent. Additionally, the '176 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. 10,152,973. The '176 patent or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on August 10, 2018. As of that date, the '097 patent had issued claiming priority to the '176 patent. On information and belief, Amazon has known of, or been willfully

blind to, the existence of the ’097 patent since no later than August 10, 2018. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the ’097 patent. Additionally, U.S. Pat. No. 8,145,489 or its published application was cited during prosecution of one or more patents assigned to a subsidiary of Amazon, including U.S. Patent No. 10,049,656. U.S. Pat. No. 8,145,489 or its published application was first cited during prosecution of one or more patents assigned to a subsidiary of Amazon on January 12, 2015. On February 23, 2016, the ’097 patent issued claiming priority to U.S. Pat. No. 8,145,489. On information and belief, Amazon has known of, or been willfully blind to, the existence of the ’097 patent since no later than February 23, 2016. On information and belief, no later than that date, Amazon has known its Alexa Products infringe the ’097 patent.

108. Amazon’s infringement has been, and continues to be knowing, intentional, and willful.

109. Amazon’s acts of infringement of the ’097 patent have caused and will continue to cause VoiceBox damages for which VoiceBox is entitled to compensation pursuant to 35 U.S.C. § 284.

110. 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, actively induced infringement of, and contributorily infringed at least one claim 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; and

f) 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 29, 2019

WILSON SONSINI GOODRICH & ROSATI
Professional Corporation

/s/ Ian R. Liston

Ian R. Liston (Delaware Bar I.D. # 5507)
222 Delaware Avenue
Suite 800
Wilmington, DE 19801
iliston@wsgr.com
Telephone: (302) 304-7600
Facsimile: (866) 974-7329

Edward G. Poplawski, *pro hac vice pending*
Erik J. Carlson, *pro hac vice pending*
Ryan S. Benyamin, *pro hac vice pending*
633 West Fifth Street, Suite 1550
Los Angeles, CA 90071-2027
epoplawski@wsgr.com
ecarlson@wsgr.com
rbenyamin@wsgr.com
Telephone: (323) 210-2900
Facsimile: (866) 974-7329

James C. Yoon, *pro hac vice pending*
Ryan Smith, *pro hac vice pending*
650 Page Mill Road
Palo Alto, CA 94304-1050
jyoon@wsgr.com
rsmith@wsgr.com
Telephone: (650) 493-9300
Facsimile: (650) 493-6811

Attorneys for VB Assets, LLC

EXHIBIT A

Sign in Join now

Philippe Di Cristo

Speech and Voice Technology Leader at Amazon

Bellevue, Washington

Computer Software

6 people have recommended Philippe

Amazon

VoiceBox Technologies

Aix-Marseille University

500+ connections

View Philippe Di Cristo's full profile. It's free!

Your colleagues, classmates, and 500 million other professionals are on LinkedIn.

Join LinkedIn

Summary

Avid follower of "conversational computing" leveraging emerging AI technologies and the ever growing accessibility to people data and behaviors. Passionate about technological innovation moving the human-machine interaction forward.

32% Technical Leader and Manager, 25% Innovator, 23% Scientist, 17% Developer, 3% Still-trying-to-figure-it-out

Sign in Join now

experiences leading to pioneering intellectual property.

Come join me build the ultimate Personal Assistant experience.
I am looking for passionate scientists in Natural Language Understanding and Machine Learning. Your work directly improves Alexa customers by making her smarter everyday. We work hard, have fun, and of course make history. What are you waiting for?

Experience

Image for Speech and

Speech and Voice Technology Leader

Amazon

October 2016 – Present • 2 years 10 months

Seattle, WA

I am building a team taking to take broad responsibilities in building the ultimate Personal Assistant experiences.

I am looking for ML scientists and Computational Linguists.

Image for Chief Scientis

Chief Scientist, Vice-President Speech and NLU R&D

VoiceBox Technologies

July 2003 – September 2016 • 13 years 3 months

Bellevue, WA

I joined the company in its infant stage as employee number 9. I cover broad responsibilities including strong technology lead, top-level management and architect/SE.

Some of my functions include leading an international team of over 100 engineers and scientists designing and implementing new Speech Recognition and Natural Language Understanding algorithms to enhance company's IP, researching and identifying technology trends to define mission critical strategic and tactical roadmap and patents, researching and evaluating worldwide potential technology partners, communicating company's technology IP to investors, executives partners, customers and stakeholders as well as driving competitor analysis and managing Speech partner relations.

Sign in Join now

for Lexus Enform/Entune, Samsung S-Voice (now on all S6 and S7 phones), as well as worldwide embedded car products (Chrysler, Renault, and more).

Designed and implemented in 2006, VoiceBox “Cybermind”, the Amazon Echo-like system. Look for it on YouTube! (http://bit.ly/cybermind). In 2013, IEEE ranked VoiceBox Technologies #13 in Patent Power for the Computer Software industry.

Report to executives and work closely with finance and sales and marketing department to ensure cohesive execution of the company’s plan

Result from IEEE that says it all: "Among the world’s most Creative and Impactful Technology IP"

Image for Director Core

Director Core Technologies

Conversay

2000 – 2003 • 3 years

Redmond WA

As Principal Speech Scientist and Architect, I lead the embedded Speech Recognition and Text-to-Speech (TTS) research and development. I am responsible not only for the design and management of the entire system architecture but also for improving the performances of “CASSI” (Conversay ASR and Speech Synthesis Interface) by proposing new algorithms and long-term research projects. I manage a domestic and international scientist team.
I work directly with C level and monthly report to the Board of Directors.

Image for Sr. Softwar e

Sr. Software Engineer

Eloquent Technologies (now part of Nuance)

1999 – 2000 • 1 year

Ithaca, New York Area

As main Speech Scientist, I am responsible for the enhancement and maintenance of the core text to speech synthesis engine. I designed a new architecture and optimized the Delta language compiler and interpreter (a proprietary programming language of Eloquent Technology, designed for speech synthesis). I also performed analysis and improved French prosody rules leveraging linguistic knowledge.

Sign in Join now

Image for Research

Research Engineer, NLP Lead

Scientific National Research Center (CNRS) - Speech and Language Lab (Laboratoire Parole et Langage)

1994 – 1998 • 4 years

Marseille Area, France

As NLP Lead within the Language Research Engineering (LRE) projects Multext (MultiLingual Text Tools and Corpora) which encompassed initially 7 European countries (France, Spain, Netherlands, Germany, Italy and Switzerland), I contributed to the Development of standards and specifications for the encoding and the processing of linguistic corpora and, the production of tools and resources for natural language processing (NLP) in a multilingual environment. Multext was the largest project funded in the Commission of European Communities Linguistic Research and Engineering Program; it got extended to include eastern Europe countries and partners (Multext-East)

Education

Image for Aix-Marseille

Aix-Marseille University

Ph.D, Speech & Computer Science

1993 – 1997

Volunteering Experience

Image for Honors selectio

Honors selection committee members of the Innovation FABA 2016

French American Chamber of Commerce

May 2016

Science and Technology

As an active member in the community, I am supporting the French American Chamber of Commerce Business Innovation Award (FABA) 2016I

http://faccpnw.wix.com/thefabapnw#!innovation-award/xjs0v

Skills & Endorsements

Sign in Join now

Languages

English
Full professional proficiency

French
Native or bilingual proficiency

Spanish
Limited working proficiency

Recommendations

A preview of what LinkedIn members have to say about Philippe:

" Philippe is "captain oh my captain"! I have been working under his supervision for close to 10 years. Thanks to him, during the whole time I have been always feeling the passion for work, looking forward to new challenges, and feel lucky to work in such a great (yet with a lot pressure) work environment. He does not only have a very strong technical background, programming skills, clear/long-term vision from both technology and application point of views, but also a great personality. He brings laughter to everywhere. It is really an honor and pleasure to work with him. I am sure he will succeed wherever he is going.

" Philippe is a great leader. Under his supervision, we were able to achieve great success on very difficult tasks. Philippe understands how engineering and research works, and he is also able to manage difficult requests from demanding customers. While he was always focused on quality and speed of delivery, he was also very respectful of personal boundaries and understands that life extends beyond work. He is a real pleasure to work with and I highly recommend him!

Join LinkedIn to see who recommended Philippe

Sign in Join now

Groups

Image **The Language Technology Group**

Image **Automatic Speech Recognition**

Image **Natural Language Processing People**

Image **Speech Technology**

Image **Computational Linguistics**

Image **Speech Analytics**

Show 2 more groups

View Philippe Di Cristo's full profile to...

- See who you know in common
- Get introduced
- Contact **Philippe Di Cristo** directly

Join LinkedIn

People also viewed

Sign in Join now

Noah B Arnett
Technology Office Assistant at City of Dover

Janna Wagner
Student at Kent State University at Trumbull

Kayla Murphy
Student at Kent State University at Trumbull

Yang Mi
Student at University of Washington

Grant Peterson
Student Majoring in Supply Chain Management at University of Tennessee, Knoxville

Nathan Yonker
High School Diploma at Ohio Virtual Academy and student of Kent state Trumbull

Devin Smeallie

Krishel Arrell
Client and communication focused individual invested in the promotion of positivity within my environment.

Rajiv Rai
Coach

Search for other people you know

First name Last name

Sign in Join now

Learn new skills with online courses

AI **AI Accountability Essential Training**

Programming **Programming Foundations: Test-Driven Development (2013)**

Blockchain: **Blockchain: Learning Solidity**

View all online courses

Philippe's public profile badge

Include this LinkedIn profile on other websites

Phili ppe **Philippe Di Cristo**
Speech and Voice Technology Leader at Amazon

Speech and Voice Technology Leader at Amazon

Aix-Marseille University

View profile

Sign in Join now

View similar profiles

Moha med Kabou

Mohamed Kaboud

Technology Lead at Expedia, Inc.

Emma Hardm an

Emma Hardman

Accessibility Technology Lead at Expedia Group

Avina sh Gupta

Avinash Gupta Kolluri

Technology Lead / Cloud Solutions Design Engineer

Dan Corlis s

Dan Corliss

Technology Lead AR/VR at Facebook

Anil Kumar

Anil Kumar

Technology Lead at Expedia

Rekha Nanda

Rekha Nanda

Technology Leader , International Technology at Amazon

Breck Morris on

Breck Morrison

Technology Leader at Amazon

Sumal atha Koppu

Sumalatha Koppula

Technology Lead at Expedia Inc | AWS Certified Solutions Architect

Tony Wu

Tony Wu

Hadoop Administrator at Expedia, Inc.

Velsa my Raja

Velsamy Raja

Technology Lead at Expedia, Inc.

Sign in Join now

© 2019

About

User Agreement

Privacy Policy

Cookie Policy

Copyright Policy

Brand Policy

Guest Controls

Community Guidelines

Language

EXHIBIT B

voicebox
1110 112th Ave NE Suite 500
Bellevue, Washington 98004

Jeff Bezos
Amazon.com Inc
410 Terry Avenue North
Seattle, Washington
98109

January 17, 2017

Dear Jeff,

This is Mike Kennewick, CEO of Voicebox Technologies in Bellevue, WA. We, along with Nuance Communications, are the leader in Speech and NLU services for the automotive marketplace. We also developed an NLU for Samsung for use in their smart phones -- including their upcoming Personal Assistant, Bixby. I met you briefly when you were just starting Amazon, you visited my Document Management Company, Saros. I have followed and admired your amazing success ever since.

The reason I am writing is that your Voice and Advanced NUI Shopping team recently hosted a targeted, invite-only recruiting dinner for over 100 of our current and former employees. All of our employees have signed a non-compete and non-solicitation agreement, so we are very concerned about Amazon's recruitment of them. Compounding the matter is that a former Voicebox Chief Speech Scientist & Vice President, Embedded Speech Technologies, who is in violation of both his non-compete and non-solicitation agreement, was heavily involved in and listed as a speaker at the event. Also, I'd like to bring to your attention the fact that your team came to our office in 2011 to discuss a potential acquisition and we disclosed details about a Smart Speaker/Personal Assistant we created. The following link is a demonstration of our Smart Speaker: King 5 News https://youtu.be/DDcRyPnvWhw. As you can see, the Echo is remarkably similar in design and functionality to our 2005 product.

Our lawyers are in discussion with your lawyers about the recruiting issue and your law firm, Summit Law Group, asked us whether there is a business solution to the problem. The reason I am writing is to see if we can in fact find such a solution.

As it turns out, we are currently considering the sale of Voicebox to another party. It occurs to me that perhaps we should also be also talking to Amazon about such an acquisition:

1. We have 150 people in Bellevue, who are experts in Speech and NLU, whose talent your team is obviously interested in.
2. We have 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.
3. We have extensive experience in the automotive market and we are the market leader in providing Natural Language Understanding to car companies including Toyota, Renault, Mazda, Subaru and Chrysler.

Your team's interest in our engineers and your recent announcements about the connected car makes me think that there may be strong synergies between our companies. I am particularly struck by the public statements you've made about innovation as our company is driven by innovation; we were the first to develop a personal assistant in a smart speaker and we were the first to develop a personal assistant in a connected car. In summary, we are an innovator right in your own back yard and we have people, IP and capabilities that I believe may make more sense with Amazon than the other alternative we are considering. I am sincere in my belief that the combination of our capabilities and yours can make a big difference against competitors like Google, Apple and Alibaba. I hope you will consider my offer in the spirit it is intended.

Please let me know if you are interested in a short phone call or face to face meeting to discuss this possibility.

Sincerely Yours,

Mike Kennewick

EXHIBIT C

US008073681B2

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

(10) **Patent No.: US 8,073,681 B2**
(45) **Date of Patent: Dec. 6, 2011**

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

(75) 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)

(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 1217 days.

(21) Appl. No.: **11/580,926**

(22) Filed: **Oct. 16, 2006**

(65) **Prior Publication Data**

US 2008/0091406 A1 Apr. 17, 2008

(51) **Int. Cl.**
G06F 17/27 (2006.01)
G10L 21/00 (2006.01)
G10L 11/00 (2006.01)

(52) **U.S. Cl.** **704/9**; 704/270; 704/275

(58) **Field of Classification Search** 704/270–275, 704/9

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
5,274,560	A	12/1993	LaRue	364/444
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
5,488,652	A *	1/1996	Bielby et al.	379/88.03
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

(Continued)

FOREIGN PATENT DOCUMENTS

EP 1 320 043 A2 6/2003

(Continued)

OTHER PUBLICATIONS

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.

(Continued)

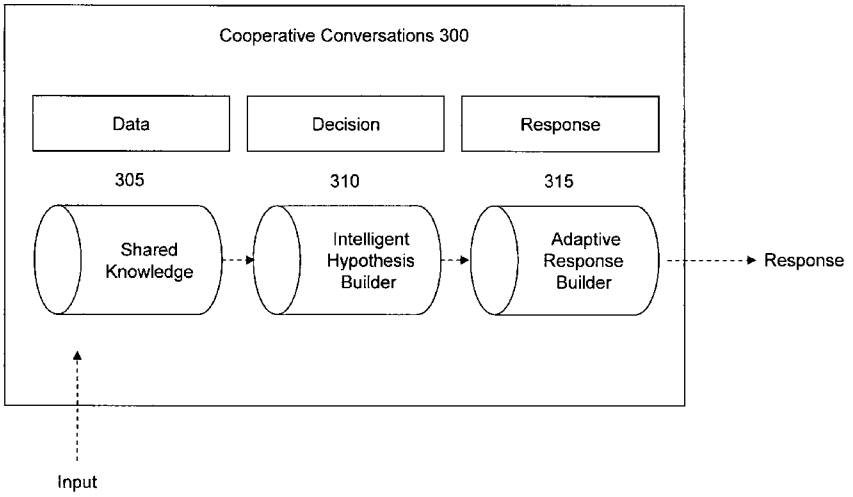
Primary Examiner — Eric Yen

(74) *Attorney, Agent, or Firm* — Pillsbury Winthrop Shaw Pittman 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.

42 Claims, 3 Drawing Sheets



U.S. PATENT DOCUMENTS

5,539,744	A		7/1996	Chu et al. 370/60
5,557,667	A		9/1996	Bruno et al. 379/201
5,563,937	A		10/1996	Bruno et al. 379/201
5,590,039	A		12/1996	Ikeda et al. 395/759
5,617,407	A		4/1997	Bareis 369/275.3
5,633,922	A		5/1997	August et al. 379/220
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
5,742,763	A		4/1998	Jones 395/200.3
5,748,841	A	*	5/1998	Morin et al. 704/257
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
5,774,859	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
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
5,832,221	A		11/1998	Jones 375/200.36
5,839,107	A		11/1998	Gupta et al. 704/270
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
5,895,464	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
5,902,347	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 et al. 707/1
5,926,784	A		7/1999	Richardson et al. 704/9
5,933,822	A		8/1999	Braden-Harder et al. 707/5
5,953,393	A		9/1999	Culbreth et al. 379/88.25
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
5,987,404	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
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
6,044,347	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
6,078,886	A		6/2000	Dragosh et al. 704/270
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
6,134,235	A		10/2000	Goldman et al. 370/352
6,144,667	A		11/2000	Doshi et al. 370/401
6,144,938	A		11/2000	Surace et al. 704/257
6,154,526	A	*	11/2000	Dahlke et al. 379/88.03
6,160,883	A		12/2000	Jackson et al. 379/230
6,167,377	A		12/2000	Gillick et al. 704/240
6,173,266	B1		1/2001	Marx et al. 704/270
6,173,279	B1		1/2001	Levin et al. 707/5
6,175,858	B1		1/2001	Bulfer et al. 709/206
6,185,535	B1		2/2001	Hedin et al. 704/270
6,192,110	B1		2/2001	Abella et al. 379/88.01
6,192,338	B1		2/2001	Haszto et al. 704/257
6,195,634	B1		2/2001	Dudemaine et al. 704/231
6,195,651	B1		2/2001	Handel et al. 707/2
6,208,964	B1		3/2001	Sabourin 704/244
6,208,972	B1		3/2001	Grant et al. 704/275
6,219,346	B1		4/2001	Maxemchuk 370/338
6,219,643	B1		4/2001	Cohen et al. 704/257
6,226,612	B1		5/2001	Srenger et al. 704/256
6,233,556	B1		5/2001	Teunen et al. 704/250
6,233,559	B1		5/2001	Balakrishnan 704/275
6,246,981	B1		6/2001	Papineni et al. 704/235
6,269,336	B1		7/2001	Ladd et al. 704/270
6,272,455	B1		8/2001	Hoshen et al. 704/1
6,278,968	B1	*	8/2001	Franz et al. 704/3
6,288,319	B1		9/2001	Catona 84/609
6,292,767	B1		9/2001	Jackson et al. 704/1
6,308,151	B1		10/2001	Smith 704/235
6,314,402	B1		11/2001	Monaco et al. 704/275
6,366,882	B1		4/2002	Bijl et al. 704/235
6,366,886	B1		4/2002	Dragosh et al. 704/270.1
6,374,214	B1	*	4/2002	Friedland et al. 704/235
6,377,913	B1		4/2002	Coffman et al. 704/8
6,381,535	B1		4/2002	Durocher et al. 701/202
6,385,596	B1		5/2002	Wiser et al. 705/51
6,385,646	B1		5/2002	Brown et al. 709/217
6,393,428	B1		5/2002	Miller et al. 707/102
6,397,181	B1		5/2002	Li et al. 704/256
6,404,878	B1		6/2002	Jackson et al. 379/221.01
6,405,170	B1		6/2002	Phillips et al. 704/270
6,408,272	B1		6/2002	White et al. 704/270.1
6,411,810	B1		6/2002	Maxemchuk 455/453
6,415,257	B1		7/2002	Junqua et al. 704/275
6,418,210	B1		7/2002	Sayko 379/142.15
6,420,975	B1		7/2002	DeLine et al. 340/815.4
6,429,813	B2		8/2002	Feigen 342/357.13
6,430,285	B1		8/2002	Bauer et al. 379/265.01
6,430,531	B1		8/2002	Polish 704/257
6,434,523	B1		8/2002	Monaco 704/257
6,434,524	B1		8/2002	Weber 704/257
6,442,522	B1		8/2002	Carberry et al. 704/257
6,446,114	B1		9/2002	Bulfer et al. 709/206
6,453,153	B1		9/2002	Bowker et al. 455/67.4
6,453,292	B2		9/2002	Ramaswamy et al. 704/235
6,456,711	B1		9/2002	Cheung et al. 379/265.09
6,466,654	B1		10/2002	Cooper et al. 379/88.01
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
6,513,006	B2		1/2003	Howard et al. 704/257
6,522,746	B1		2/2003	Marchok et al. 379/406.03
6,523,061	B1		2/2003	Halverson et al. 709/202
6,532,444	B1		3/2003	Weber 704/257
6,539,348	B1		3/2003	Bond et al. 704/9
6,549,629	B2		4/2003	Finn et al. 381/92
6,553,372	B1		4/2003	Brassell et al. 707/5
6,556,970	B1		4/2003	Sasaki et al. 704/257
6,556,973	B1		4/2003	Lewin 704/277
6,560,576	B1		5/2003	Cohen et al. 704/270
6,567,778	B1		5/2003	Chao Chang et al. 704/257
6,567,797	B1		5/2003	Schuetze et al. 707/2
6,570,555	B1		5/2003	Prevost et al. 345/156
6,570,964	B1		5/2003	Murveit et al. 379/67.1
6,574,597	B1		6/2003	Mohri et al. 704/251
6,574,624	B1		6/2003	Johnson et al. 707/5
6,581,103	B1		6/2003	Dengler 709/231
6,587,858	B1		7/2003	Strazza 707/102
6,594,257	B1		7/2003	Doshi et al. 370/352
6,598,018	B1		7/2003	Junqua 704/251
6,604,075	B1		8/2003	Brown et al. 704/270.1
6,604,077	B2		8/2003	Dragosh et al. 704/270.1
6,606,598	B1		8/2003	Holthouse et al. 704/275
6,611,692	B2		8/2003	Raffel et al. 455/552
6,614,773	B1		9/2003	Maxemchuk 370/337
6,615,172	B1		9/2003	Bennett et al. 704/257
6,622,119	B1		9/2003	Ramaswamy et al. 704/9
6,629,066	B1		9/2003	Jackson et al. 704/9
6,631,346	B1		10/2003	Karaorman et al. 704/9
6,643,620	B1		11/2003	Contolini et al. 704/270
6,650,747	B1		11/2003	Bala et al. 379/265.06
6,658,388	B1		12/2003	Kleindienst et al. 704/275
6,678,680	B1		1/2004	Woo 707/6
6,681,206	B1		1/2004	Gorin et al. 704/243
6,691,151	B1		2/2004	Cheyer et al. 709/202
6,701,294	B1		3/2004	Ball et al. 704/257

6,704,708	B1	3/2004	Pickering	704/235
6,708,150	B1	3/2004	Hirayama et al.	704/243
6,721,001	B1	4/2004	Berstis	348/231.3
6,721,706	B1	4/2004	Strubbe et al.	704/275
6,735,592	B1	5/2004	Neumann et al.	707/101
6,741,931	B1	5/2004	Kohut et al.	701/209
6,742,021	B1	5/2004	Halverson et al.	709/218
6,751,591	B1 *	6/2004	Gorin et al.	704/257
6,751,612	B1	6/2004	Schuetze et al.	707/4
6,757,718	B1	6/2004	Halverson et al.	709/218
6,795,808	B1	9/2004	Strubbe et al.	704/275
6,801,604	B2	10/2004	Maes et al.	379/88.17
6,801,893	B1	10/2004	Backfried et al.	704/257
6,829,603	B1 *	12/2004	Wolf et al.	1/1
6,832,230	B1	12/2004	Zilliacus et al.	707/203
6,833,848	B1	12/2004	Wolff et al.	345/719
6,856,990	B2	2/2005	Barile et al.	707/10
6,865,481	B2	3/2005	Kawazoe et al.	701/211
6,877,134	B1	4/2005	Fuller et al.	715/500.1
6,901,366	B1	5/2005	Kuhn et al.	704/275
6,912,498	B2	6/2005	Stevens et al.	704/235
6,937,977	B2	8/2005	Gerson	704/201
6,944,594	B2	9/2005	Busayapongchai et al.	704/275
6,950,821	B2	9/2005	Faybishenko et al.	707/10
6,954,755	B2	10/2005	Reisman	707/10
6,973,387	B2	12/2005	Masclet et al.	701/211
6,975,993	B1	12/2005	Keiller	704/275
6,980,092	B2	12/2005	Turnbull et al.	340/425.5
6,990,513	B2	1/2006	Belfiore et al.	709/203
6,996,531	B2	2/2006	Korall et al.	704/270
7,003,463	B1	2/2006	Maes et al.	704/270.1
7,016,849	B2	3/2006	Arnold et al.	704/275
7,020,609	B2	3/2006	Thrift et al.	704/270.1
7,024,364	B2	4/2006	Guerra et al.	704/270
7,027,975	B1	4/2006	Pazandak et al.	704/9
7,058,890	B2	6/2006	George et al.	715/728
7,062,488	B1	6/2006	Reisman	707/8
7,082,469	B2	7/2006	Gold et al.	709/231
7,092,928	B1	8/2006	Elad et al.	706/60
7,127,400	B2	10/2006	Koch	704/270.1
7,136,875	B2	11/2006	Anderson et al.	707/104.1
7,137,126	B1	11/2006	Coffman et al.	719/328
7,146,319	B2	12/2006	Hunt	704/254
7,197,069	B2	3/2007	Agazzi et al.	375/233
7,203,644	B2	4/2007	Anderson et al.	704/246
7,228,276	B2	6/2007	Omote et al.	704/243
7,231,343	B1	6/2007	Treadgold et al.	704/9
7,236,923	B1	6/2007	Gupta	704/9
7,277,854	B2	10/2007	Bennett et al.	704/257
7,289,606	B2	10/2007	Sibal et al.	379/52
7,301,093	B2	11/2007	Sater et al.	84/615
7,340,040	B1	3/2008	Saylor et al.	379/67.1
7,366,669	B2	4/2008	Nishitani et al.	704/256
7,376,645	B2	5/2008	Bernard	707/3
7,386,443	B1	6/2008	Parthasarathy et al.	704/201
7,398,209	B2	7/2008	Kennewick et al.	704/255
7,415,414	B2	8/2008	Azara et al.	704/270
7,424,431	B2	9/2008	Greene et al.	704/270
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
7,487,088	B1 *	2/2009	Gorin et al.	704/240
7,493,259	B2	2/2009	Jones et al.	704/257
7,493,559	B1	2/2009	Wolff et al.	715/727
7,502,738	B2	3/2009	Kennewick et al.	704/257
7,536,297	B2	5/2009	Byrd et al.	704/10
7,536,374	B2 *	5/2009	Au	706/55
7,558,730	B2	7/2009	Davis et al.	704/235
7,620,549	B2	11/2009	Di Cristo et al.	704/257
7,634,409	B2	12/2009	Kennewick et al.	704/257
7,640,160	B2	12/2009	Di Cristo et al.	704/257
7,676,365	B2	3/2010	Hwang et al.	704/240
7,693,720	B2	4/2010	Kennewick et al.	704/275
7,788,084	B2 *	8/2010	Brun et al.	704/7
7,809,570	B2	10/2010	Kennewick et al.	704/257
7,818,176	B2	10/2010	Freeman et al.	704/270
7,831,433	B1	11/2010	Belvin et al.	704/275
7,873,523	B2	1/2011	Potter et al.	704/275
7,917,367	B2	3/2011	Di Cristo et al.	704/270.1
7,949,529	B2	5/2011	Weider et al.	704/270
7,983,917	B2	7/2011	Kennewick et al.	704/257
8,015,006	B2	9/2011	Kennewick et al.	704/236
2001/0041980	A1	11/2001	Howard et al.	704/270
2001/0049601	A1 *	12/2001	Kroeker et al.	704/254
2002/0035501	A1	3/2002	Handel et al.	705/10
2002/0049805	A1	4/2002	Yamada et al.	709/202
2002/0065568	A1	5/2002	Silfvast et al.	700/94
2002/0069059	A1	6/2002	Smith	704/257
2002/0082911	A1	6/2002	Dunn et al.	705/14
2002/0087525	A1	7/2002	Abbott et al.	707/3
2002/0120609	A1	8/2002	Lang et al.	707/1
2002/0124050	A1	9/2002	Middeljans	709/203
2002/0138248	A1	9/2002	Corston-Oliver et al.	704/1
2002/0143535	A1	10/2002	Kist et al.	704/251
2002/0188602	A1	12/2002	Stubler et al.	707/3
2002/0198714	A1	12/2002	Zhou	704/252
2003/0014261	A1	1/2003	Kageyama	704/275
2003/0046346	A1	3/2003	Mumick et al.	709/205
2003/0064709	A1	4/2003	Gailey et al.	455/412
2003/0088421	A1	5/2003	Maes et al.	704/270.1
2003/0097249	A1 *	5/2003	Walker et al.	704/1
2003/0112267	A1	6/2003	Belrose	345/728
2003/0120493	A1	6/2003	Gupta	704/270.1
2003/0135488	A1	7/2003	Amir et al.	707/3
2003/0144846	A1	7/2003	Denenberg et al.	704/277
2003/0182132	A1	9/2003	Niemoeller	704/275
2003/0212550	A1	11/2003	Ubale	704/215
2003/0236664	A1	12/2003	Sharma	704/251
2004/0006475	A1	1/2004	Ehlen et al.	704/270.1
2004/0025115	A1	2/2004	Sienel et al.	715/513
2004/0044516	A1	3/2004	Kennewick et al.	704/5
2004/0166832	A1	8/2004	Portman et al.	455/412.1
2004/0167771	A1 *	8/2004	Duan et al.	704/10
2004/0193408	A1	9/2004	Hunt	704/209
2004/0193420	A1	9/2004	Kennewick et al.	704/257
2004/0199375	A1	10/2004	Ehsani et al.	704/4
2004/0243417	A9	12/2004	Pitts, III et al.	704/276
2005/0015256	A1 *	1/2005	Kargman	704/272
2005/0021334	A1	1/2005	Iwahashi	704/240
2005/0021826	A1	1/2005	Kumar	709/232
2005/0033574	A1	2/2005	Kim et al.	704/251
2005/0043940	A1	2/2005	Elder	704/9
2005/0114116	A1	5/2005	Fiedler	704/201
2005/0137850	A1	6/2005	Odell	704/4
2005/0246174	A1	11/2005	DeGolia	704/270
2006/0206310	A1	9/2006	Ravikumar et al.	704/9
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
2007/0050191	A1	3/2007	Weider et al.	704/275
2007/0055525	A1	3/2007	Kennewick et al.	704/257
2007/0118357	A1 *	5/2007	Kasravi et al.	704/10
2007/0179778	A1	8/2007	Gong et al.	704/9
2007/0186165	A1	8/2007	Maislos et al.	715/728
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
2008/0091406	A1	4/2008	Baldwin et al.	704/4
2008/0103761	A1	5/2008	Printz et al.	704/9
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
2008/0177530	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
2009/0117885	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
2009/0299745	A1	12/2009	Kennewick et al.	704/257
2010/0023320	A1	1/2010	Di Cristo et al.	704/9

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
2010/0145700	A1	6/2010	Kennewick et al.	704/ 257
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
2011/0112827	A1	5/2011	Kennewick et al.	704/9
2011/0112921	A1	5/2011	Kennewick et al.	705/26.1
2011/0131036	A1	6/2011	Di Cristo et al.	704/9
2011/0131045	A1	6/2011	Cristo et al.	704/249

FOREIGN PATENT DOCUMENTS

EP	1 646 037	4/2006
WO	WO 99/46763	9/1999
WO	WO 00/21232	4/2000
WO	WO 00/46792	8/2000
WO	WO 2004/072954	8/2001
WO	WO 01/78065	10/2001
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

Cheyer 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, [no date], 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, [no date], 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.

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.

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

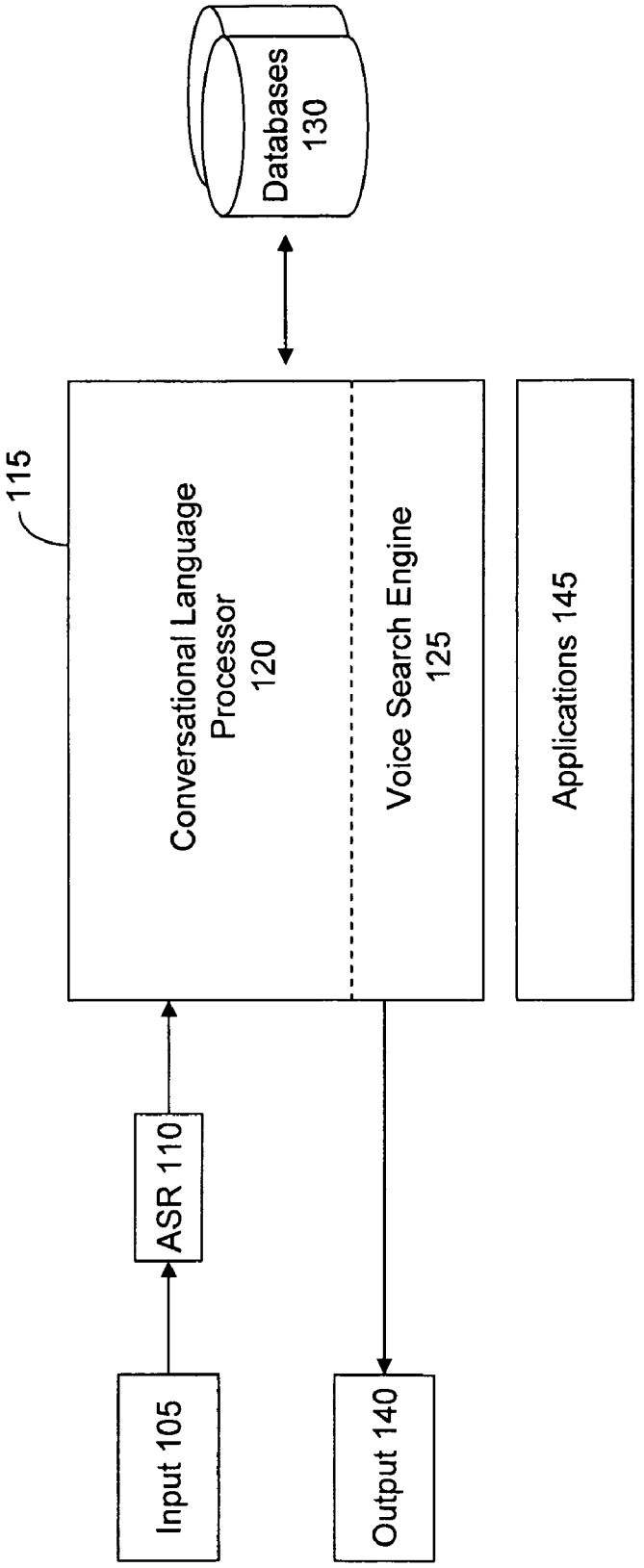

Databases
130
115
Conversational Language
Processor
120
Voice Search Engine
125
Applications 145
ASR 110
Input 105
Output 140

FIG. 1

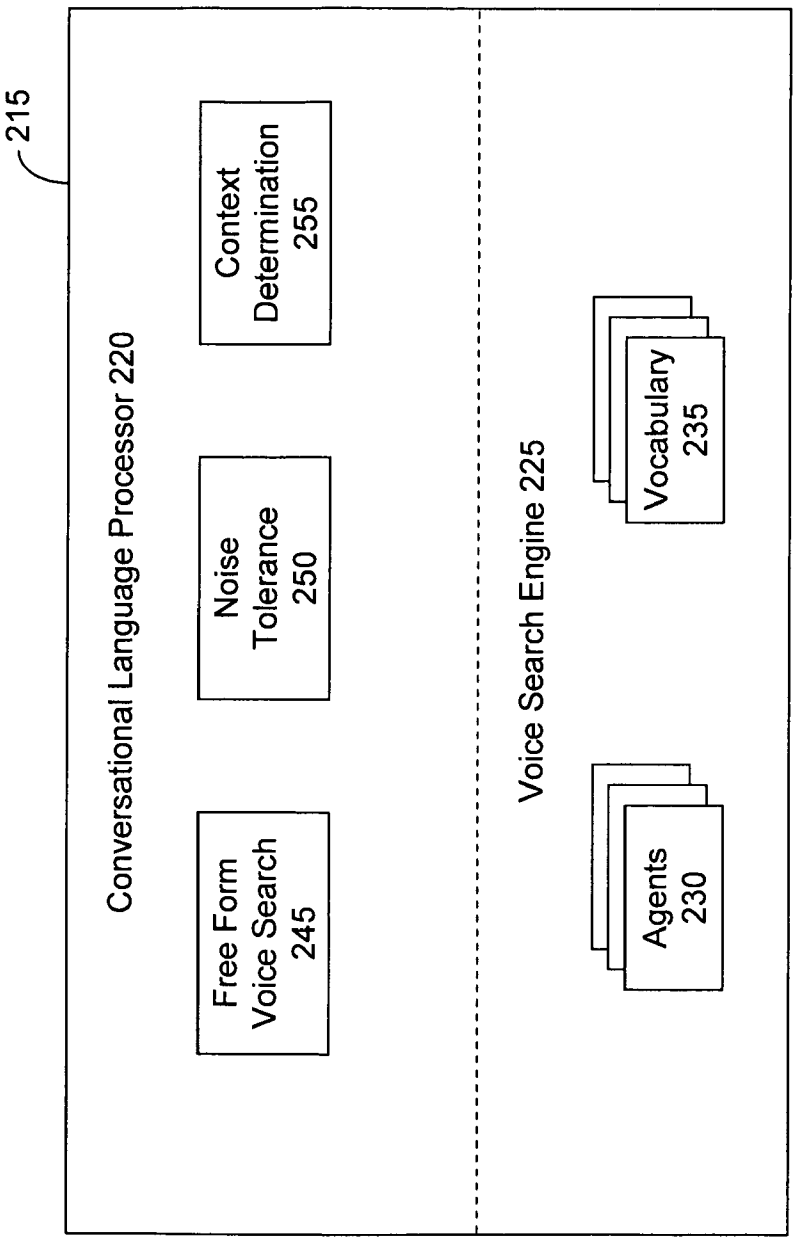

215
Conversational Language Processor 220
Context Determination 255
Noise Tolerance 250
Free Form Voice Search 245
Voice Search Engine 225
Vocabulary 235
Agents 230

FIG. 2

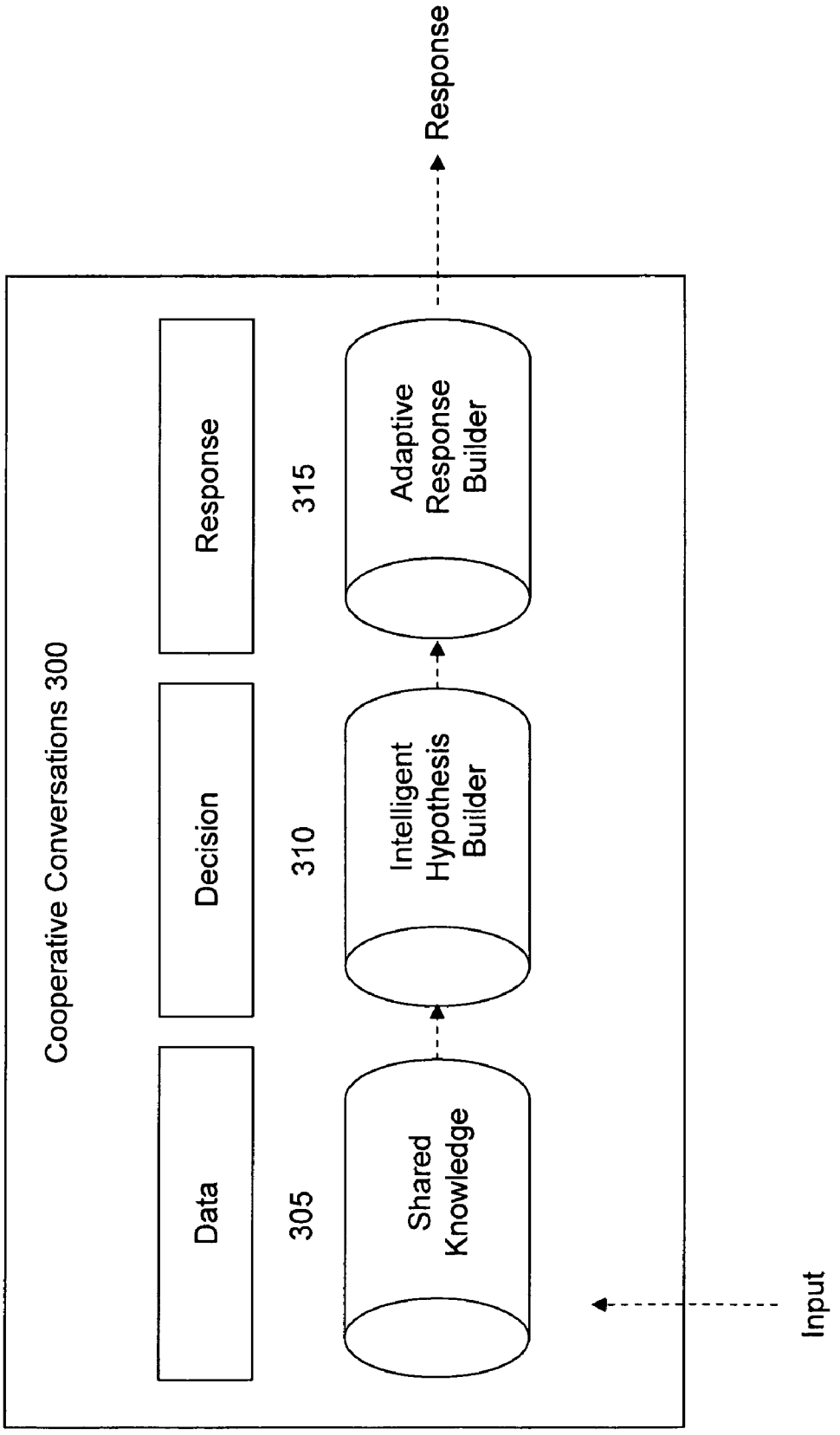

Response
Cooperative Conversations 300
Response
Decision
Data
315
310
305
Adaptive Response Builder
Intelligent Hypothesis Builder
Shared Knowledge
Input

FIG. 3

SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL VOICE USER INTERFACE

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 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 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 "wins" the competition, the winning domain may be responsible for establishing or inferring further contexts and updating short-term and long-term shared knowledge. If there is a deadlock between context domain agents, an adaptive conversational response may prompt the user to assist in disambiguating between the deadlocked agents. Moreover, the context determination process may infer intended operations and/or context based on previous utterances and/or requests, whereas existing systems consider each utterance independently, potentially making the same errors over and over again. For example, if a given interpretation turns 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 participant in a conversation with an accepted purpose or direction, with each participant contributing to the conversation for the benefit of the conversation. By taking advantage of human presumptions about utterances that humans rely upon, both as speakers and listeners, a Human-to-Machine interface may be analogous to everyday human-to-human conversation. In one implementation, the exemplary cooperative conversation model may take incoming data (shared knowledge) to inform a decision (intelligent hypothesis building), and then may refine the decision and generate a response (adaptive response building).

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

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

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). 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 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 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, 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.

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, Oreg. based on the Weather and an environmental proximity between Portland, Oreg. 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, Oreg.?").

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

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 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, 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 conver-

sation, 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		
	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 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 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?
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 d 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 cer-

tainty, which may be used to correct a course of a conversation without interrupting conversational flow. That is, even though the intelligent responses may be reasonably "sure," the intelligent responses may nonetheless sometimes be incorrect. While existing voice user interfaces tend to fail on average conversational missteps, normal human interactions may expect missteps and deal with them appropriately. Thus, responses after a misrecognition may be modeled after clarifications, rather than errors, and words may chosen in subsequent responses that move conversation forward and establish an appropriate domain to be explored with the user. For example, course correction may result in the following exchange:

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 method for providing a cooperative conversational voice user interface, comprising:

receiving an utterance at a voice input device during a current conversation with a user, wherein the utterance includes one or more words that have different meanings in different contexts;

accumulating short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;

accumulating long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;

determining an intended meaning for the utterance, wherein determining the intended meaning for the utterance includes:

identifying, at a conversational speech engine, a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge; and

establishing the intended meaning within the identified context, wherein the conversational speech engine establishes the intended meaning within the identified context to disambiguate an intent that the user had in speaking the one or more words that have the different meanings in the different contexts; and

generating a response to the utterance, wherein the conversational speech engine grammatically or syntactically adapts the response based on the intended meaning established within the identified context.

2. The method of claim **1**, wherein accumulating the short-term shared knowledge about the current conversation includes populating a short-term context stack with information about the utterance received during the current conversation.

3. The method of claim **2**, wherein accumulating the short-term shared knowledge about the current conversation further includes expiring the information about the utterance from the short-term context stack after a psychologically appropriate amount of time.

4. The method of claim **3** wherein accumulating the long-term shared knowledge about the user includes updating one or more long-term profiles associated with the user to include information about the utterance received during the current conversation and relevant data associated with the information expired from the short, term context stack.

5. The method of claim **1**, wherein determining the intended meaning for the utterance further includes:

identifying a conversational goal associated with the utterance, roles associated with the user and one or more other participants in the current conversation, and an information allocation among the user and the one or more other participants in the current conversation; and

classifying one or more of the utterance or the current conversation into a conversation type based on one or more of the identified conversational goal, the identified roles, or the identified information allocation, wherein the conversational speech engine further establishes the intended meaning based on the conversation type.

6. The method of claim **5** wherein the established intended meaning comprises a hypothesis having a degree of certainty about the intent that the user had in speaking the one or more words in the utterance.

7. The method of claim **6**, further comprising generating a preliminary interpretation of the utterance at a speech recognition engine coupled to the voice input device and the conversational speech engine, wherein the conversational speech engine assigns the degree of certainty to the hypothesis based on one or more of the conversation type, information associated with the identified context, or a confidence level associated with the preliminary interpretation generated at the speech recognition engine.

8. The method of claim **5**, wherein the conversational speech engine further grammatically or syntactically adapts the response based on the conversation type.

9. The method of claim **1**, wherein the conversational speech engine grammatically or syntactically adapts the response to influence a subsequent reply utterance that the conversational speech engine expects from the user during the current conversation.

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

generating multiple preliminary interpretations of the utterance at a speech recognition engine coupled to the voice input device and the conversational speech engine, wherein an initial interpretation of the utterance comprises one of the multiple preliminary interpretations having a highest confidence level; and

updating the short-term shared knowledge about the current conversation to remove the initial interpretation from the multiple preliminary interpretations in response to determining that the initial interpretation was incorrect, wherein the conversational speech engine determines the intended meaning based on one of the multiple preliminary interpretations having a next highest confidence level.

11. The method of claim **1**, wherein the user speaks the utterance in a multi-modal input that further includes one or more non-voice inputs relating to the utterance.

12. The method of claim 1, wherein the conversational speech engine generates the response in a multi-modal output that includes one or more non-voice outputs that relate to the utterance or one or more tasks executed to process a request identified from the intended meaning.

13. A non-transitory computer readable medium containing computer-executable instructions for providing a cooperative conversational voice user interface, the computer-executable instructions operable when executed to:

receive an utterance at a voice input device, during a current conversation with a user, wherein the utterance includes one or more words that have different meanings in different contexts;

accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received at the voice during the current conversation;

accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;

identify a context associated with the utterance, wherein a conversational speech engine identifies the context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge;

establish an intended meaning for the utterance within the identified context, wherein the conversational speech engine establishes the intended meaning within the identified context to disambiguate an intent that the user had in speaking the one or more words that have the different meanings in the different contexts; and

generate a response to the utterance, wherein the conversational speech engine grammatically or syntactically adapts the response based on the intended meaning established within the identified context.

14. The non-transitory computer readable medium of claim 13, wherein to accumulate the short-term shared knowledge about the current conversation, the computer-executable instructions are further operable when executed to populate a short-term context stack with information about the utterance received during the current conversation.

15. The non-transitory computer readable medium of claim 14, wherein to accumulate the short-term shared knowledge about the current conversation, the computer-executable instructions are further operable when executed to expire the information about the utterance from the short-term context stack after a psychologically appropriate amount of time.

16. The non-transitory computer readable medium of claim 15, wherein to accumulate the long-term shared knowledge about the user, the computer-executable instructions are further operable when executed to update one or more long-term profiles associated with the user to include information about the utterance received during the current conversation and relevant data associated with the information expired from the short term context stack.

17. The non-transitory computer readable medium of claim 13, wherein the computer-executable instructions are further operable when executed to:

identify a conversational goal associated with the utterance, roles associated with the user and one or more other participants in the current conversation, and an information allocation among the user and the one or more other participants in the current conversation; and

classify one or more of the utterance or the current conversation into a conversation type based on one or more of the identified conversational goal, the identified roles, or the identified information allocation, wherein the conversational speech engine further establishes the intended meaning based on the conversation type.

18. The non-transitory computer readable medium of claim 17, wherein the established intended meaning comprises a hypothesis having a degree of certainty about the intent that the user had in speaking the one or more words in the utterance.

19. The computer-readable medium of claim 18, wherein the computer-executable instructions are further operable when executed to generate a preliminary interpretation of the utterance at a speech recognition engine, wherein the conversational speech engine assigns the degree of certainty to the hypothesis based on one or more of the conversation type, information associated with the identified context, or a confidence level associated with the preliminary interpretation generated at the speech recognition engine.

20. The non-transitory computer readable medium of claim 17, wherein the conversational speech engine further grammatically or syntactically adapts the response based on conversation type.

21. The non-transitory computer readable medium of claim 13, wherein the conversational speech engine grammatically or syntactically adapts more the response to influence a subsequent reply utterance that the conversational speech engine expects from the user during the current conversation.

22. The non-transitory computer readable medium of claim 13, wherein the computer-executable instructions are further operable when executed to:

generate multiple preliminary interpretations of the utterance at a speech recognition engine, wherein an initial interpretation of the utterance comprises one of the multiple preliminary interpretations having a highest confidence level; and

update the short-term shared knowledge about the current conversation to remove the initial interpretation from the multiple preliminary interpretations in response to determining that the initial interpretation was incorrect, wherein the conversational speech engine identifies the context associated with the utterance and establishes the intended meaning for the utterance based on one of the multiple preliminary interpretations having a next highest confidence level.

23. The non-transitory computer readable medium of claim 13, wherein the user speaks the utterance in a multi-modal input that further includes one or more non-voice inputs relating to the utterance.

24. The non-transitory computer readable medium of claim 13, wherein the conversational speech engine generates the response in a multi-modal output that includes one or more non-voice outputs that relate to the utterance or one or more tasks executed to process a request identified from the intended meaning.

25. A system for providing a cooperative conversational voice user interface, comprising:

a voice input device configured to receive an utterance during a current conversation with a user, wherein the utterance includes one or more words that have different meanings in different contexts; and

a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to:

accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;

accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;

identify a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge;

establish an intended meaning for the utterance within the identified context to disambiguate an intent that the user had in speaking the one or more words that have the different meanings in the different contexts; and

generate a grammatically or syntactically adapted response to the utterance based on the intended meaning established within the identified context.

26. The system of claim **25**, wherein to accumulate the short-term shared knowledge about the current conversation, the one or more processors are further configured to populate a short-term context stack with information about the utterance received during the current conversation.

27. The system of claim **26**, wherein to accumulate the short-term shared knowledge about the current conversation, the one or more processors are further configured to expire the information about the utterance from the short-term context stack after a psychologically appropriate amount of time.

28. The system of claim **27**, wherein to accumulate the long-term shared knowledge about the user, the one or more processors are further configured to update one or more long-term profiles associated with the user to include information about the utterance received during the current conversation and relevant data associated with the information expired from the short term context stack.

29. The system of claim **25**, wherein the one or more processors are further configured to:

identify a conversational goal associated with the utterance, roles associated with the user and one or more other participants in the current conversation, and an information allocation among the user and the one or more other participants in the current conversation; and

classify one or more of the utterance or the current conversation into a conversation type based on one or more of the identified conversational goal, the identified roles, or the identified information allocation, wherein the one or more processors are further configured to establish the intended meaning based on the conversation type.

30. The system of claim **29** wherein the established intended meaning comprises a hypothesis having a degree of certainty about the intent that the user had in speaking the one or more words in the utterance.

31. The system of claim **30**, further comprising a speech recognition engine configured to generate a preliminary interpretation of the utterance, wherein the one or more processors are further configured to assign the degree of certainty to the hypothesis based on one or more of the conversation type, information associated with the identified context, a confidence level associated with the preliminary interpretation generated at the speech recognition engine.

32. The system of claim **29** wherein the one or more processors are further configured to generate the grammatically syntactically adapted response based on the conversation type.

33. The system of claim **25**, wherein the one or more processors are further configured to generate the grammatically or syntactically adapted response to influence a subsequent reply utterance expected from the user during the current conversation.

34. The system of claim **25**, further comprising a speech recognition engine configured to:

generate multiple preliminary interpretations of the utterance, wherein an initial interpretation of the utterance comprises one of the multiple preliminary interpretations having a highest confidence level; and

update the short-term shared knowledge about the current conversation to remove the initial interpretation from the multiple preliminary interpretations in response to determining that the initial interpretation was incorrect, wherein the one or more processors are configured to identify the context associated with the utterance and establish the intended meaning for the utterance based on one of the multiple preliminary interpretations having a next highest confidence level.

35. The system of claim **25**, wherein the user speaks the utterance in a multi-modal input that further includes one or more non-voice inputs relating to the utterance.

36. The system of claim **25**, wherein the grammatically or syntactically adapted response comprises a multi-modal output that includes one or more non-voice outputs that relate to the utterance or one or more tasks executed to process a request identified from the intended meaning.

37. A method for providing a cooperative conversational voice user interface, comprising:

receiving an utterance at a voice input device during a current conversation with a user;

accumulating short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;

accumulating long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;

determining an intended meaning for the utterance, wherein determining the intended meaning for the utterance includes:

identifying, at a conversational speech engine, a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge;

inferring additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context; and

establishing the intended meaning within the identified context based on the additional information inferred about the utterance; and

generating a response to the utterance based on the intended meaning established within the identified context.

38. The method of claim **37**, wherein the established intended meaning comprises an implicit hypothesis having a corresponding degree of certainty about an intent that the user had in speaking the utterance.

39. A non-transitory computer readable medium containing computer-executable instructions for providing a cooperative conversational voice user interface, the computer-executable instructions operable when executed to:

receive an utterance at a voice input device during a current conversation with a user;

accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;

accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;

identify a context associated with the utterance, wherein a conversational speech engine identifies the context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge;

infer additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context;

establish an intended meaning for the utterance within the identified context based on the additional information inferred about the utterance; and

generate a response to the utterance based on the intended meaning established within the identified context.

40. The non-transitory computer readable medium of claim **39**, wherein the established intended meaning comprises an implicit hypothesis having a corresponding degree of certainty about an intent that the user had in speaking the utterance.

41. A system for providing a cooperative conversational voice user interface, comprising:

a voice input device configured to receive an utterance during a current conversation with a user; and

a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to:

accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;

accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;

identify a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge;

infer additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context;

establish an intended meaning for the utterance within the identified identify a context based on the additional information inferred about the utterance; and

generate a response to the utterance based on the intended meaning established within the identified context.

42. The system of claim **41**, wherein the established intended meaning comprises an implicit hypothesis having a corresponding degree of certainty about an intent that the user had in speaking the utterance.

* * * * *

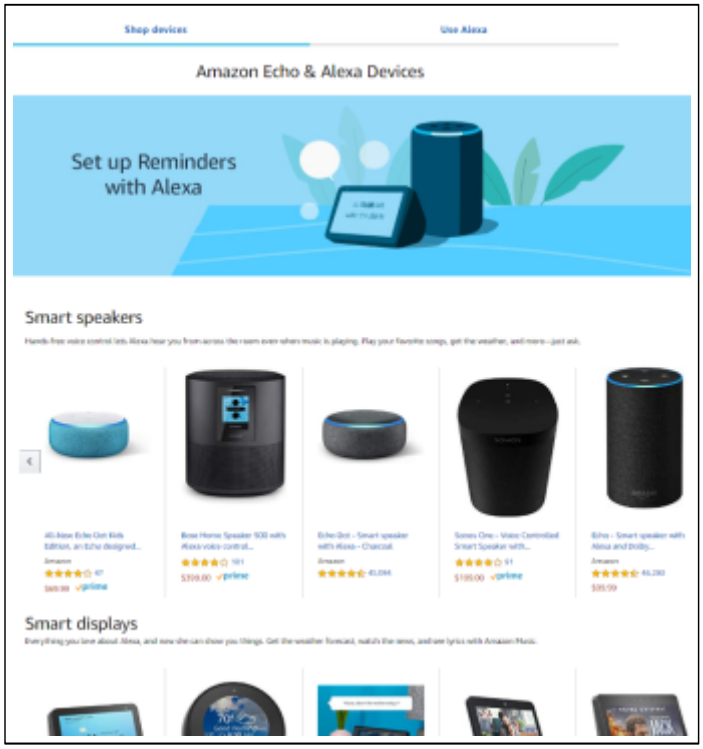
EXHIBIT D

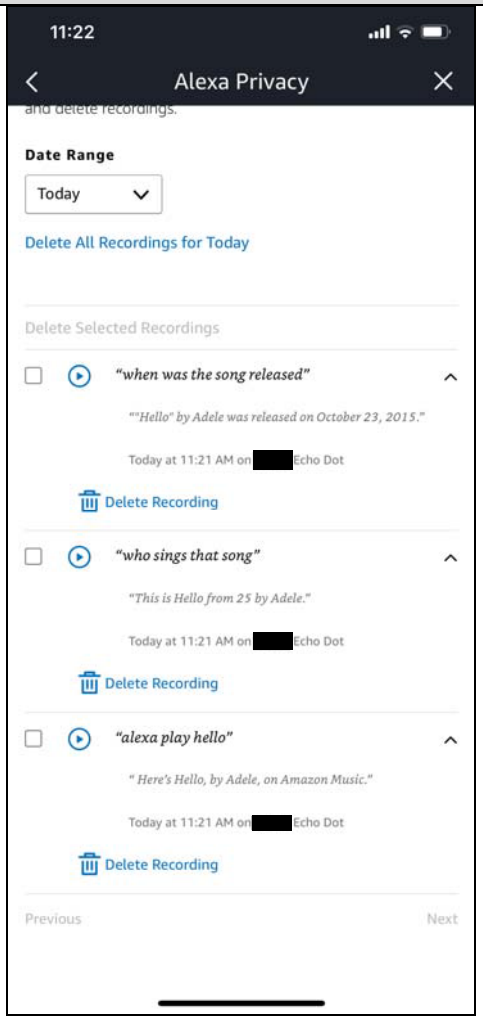
U.S. Pat. No. 8,073,681 ("the '681 patent")
Infringement Chart for Claim 41

The following chart provides exemplary infringement allegations with respect to an exemplary claim of the '681 patent based on Plaintiff's investigation to date. Plaintiff reserves the right to provide modified and/or more detailed allegations (including with additional claims, products, and evidence) in accordance with its discovery obligations in view of any applicable scheduling order. This chart uses "Alexa Products" to refer to the full scope of instrumentalities identified in the complaint. Any discussion of specific Alexa Products in the following chart is exemplary.

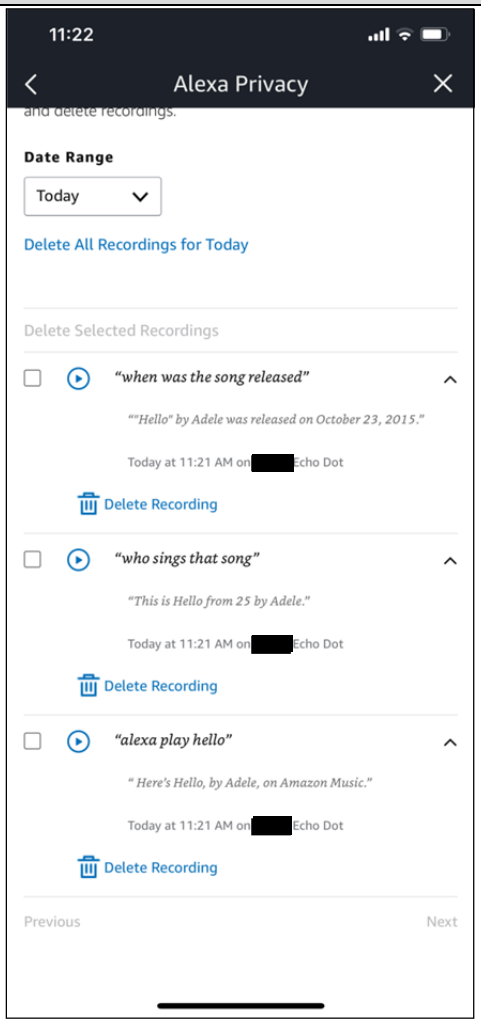
Exemplary information relevant to infringement:

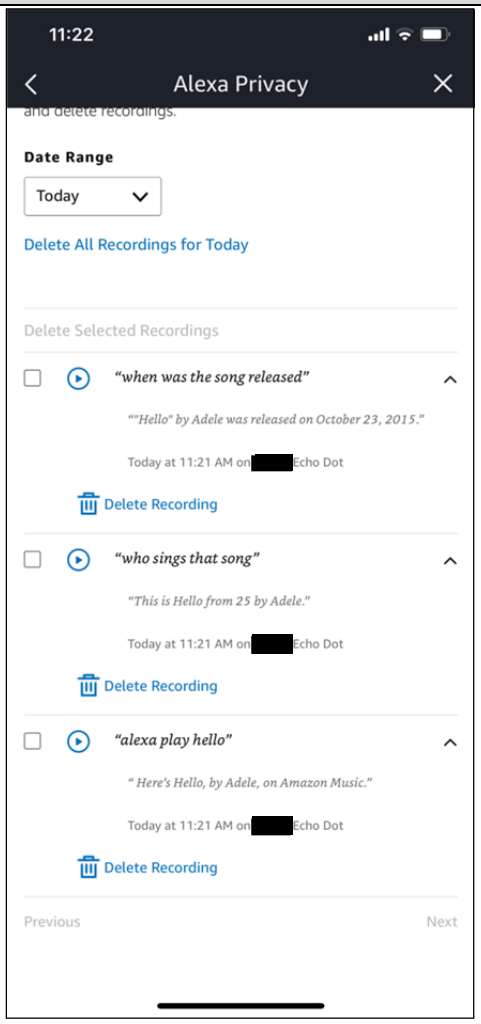
1. Larry Hardesty, *Alexa at Interspeech 2018: How Interaction Histories Can Improve Speech Understanding*, Aug. 29, 2018 post on Alexa Blogs, https://developer.amazon.com/blogs/alexa/post/37473f78-6726-4b8a-b08d-6b0d41c62753/title-tk ("*Interaction Histories*")
2. Chetan Naik et al., *Contextual Slot Carryover for Disparate Schemas*, Proc. Interspeech 2018, 596-600, *available at* https://www.isca-speech.org/archive/Interspeech_2018/pdfs/1035.pdf
3. *Amazon Echo & Alexa Devices* - Amazon Official Site, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/?ie=UTF8&node=9818047011&ref_=sv_devicesubnav_1 (last visited Jul. 22, 2019)
4. *Alexa and Alexa Device FAQs*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (last visited Jul. 22, 2019)
5. Sarah Perez, *Alexa will soon gain a memory, converse more naturally, and automatically launch skills,* TechCrunch, Apr. 26, 2018, https://techcrunch.com/2018/04/26/alexa-will-soon-gain-a-memory-converse-more-naturally-and-automatically-launch-skills/ ("*Alexa will soon gain a memory*")
6. Ruhi Sarikaya, *Making Alexa More Friction-Free*, Apr. 25, 2018 post on Alexa Blogs, https://developer.amazon.com/blogs/alexa/post/60e1f011-3236-4162-b0f6-509205d354ca/making-alexa-more-friction-free
7. Arpit Gupta, *How Alexa Is Learning to Converse More Naturally*, Aug. 27, 2018 post on Alexa Blogs, https://developer.amazon.com/blogs/alexa/post/15bf7d2a-5e5c-4d43-90ae-c2596c9cc3a6/how-alexa-is-learning-to-converse-more-naturally ("*How Alexa is Learning*")
8. Kyle Wiggers, *Amazon researchers share how AI helps Alexa answer follow-up questions*, VentureBeat, August 27, 2018, https://venturebeat.com/2018/08/27/amazon-researchers-share-how-ai-helps-alexa-answer-follow-up-questions/ ("*AI helps Alexa*")
9. *What is Conversational AI?*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/alexa-skills-kit/conversational-ai (last visited Jul. 22, 2019)
10. *Designing for Conversation – Build for dialog*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/ask-resources/guided/conversational-design-workshop#/lessons/gDcQVF9nOFOtNljrELzARIGN4HuD7LNh?_k=q4qp1a (last visited Jul. 22, 2019)
11. *Audio Hardware Configurations*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/docs/alexa-voice-service/audio-hardware-configurations.html (last visited Jul. 22, 2019)

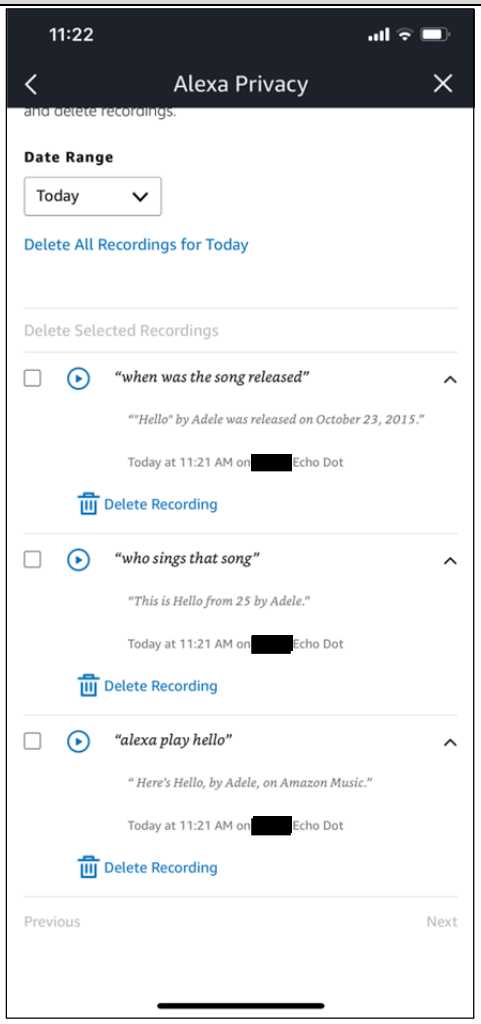
'681 Patent Claim Language	Exemplary Alexa Products
41.[pr] A system for providing a cooperative conversational voice user interface, comprising:	Alexa Products include a system for providing a cooperative conversational voice user interface. For example, Alexa Products (e.g. Echo, Echo Dot, Alexa apps, and Alexa cloud) include Alexa.  E.g. *Amazon Echo & Alexa Devices*. "With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?* Alexa Products include Amazon's Alexa cloud. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1.
[41.a] a voice input device configured to receive an utterance during a current conversation with a user; and	Alexa Products include a voice input device configured to receive an utterance during a current conversation with a user. An example is receiving an utterance shown below in a screenshot of Alexa app's "Review Voice History" page below.

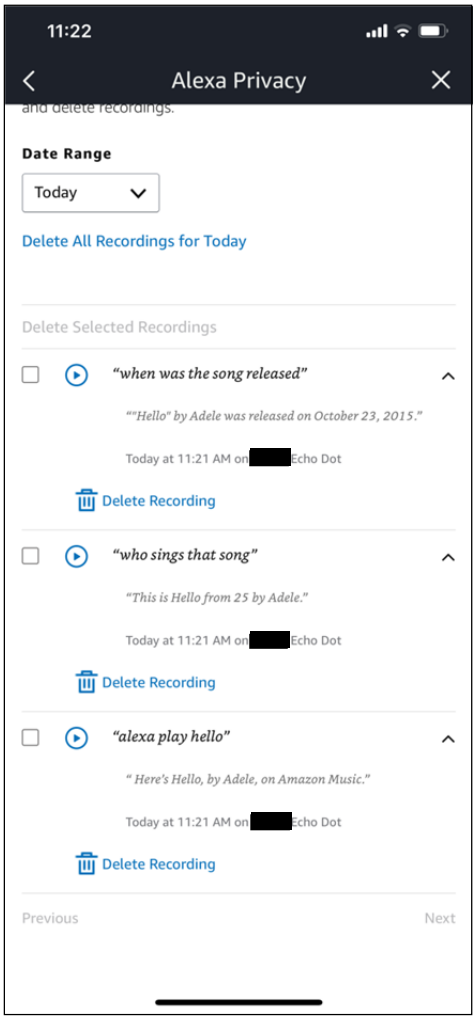
'681 Patent Claim Language	Exemplary Alexa Products
	 Below are additional examples. "Each line of the script spoken by the user is known as an utterance. The utterances are then mapped to an intent. When the user speaks one of them the associated intent will be invoked." *Designing for Conversation – Build for dialog* at Step 2. "With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?* Alexa Products include Amazon's Alexa cloud. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1.
[41.b] a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to:	Alexa Products include a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to perform the operations recited in the claim. For example, Alexa Products include Amazon's Alexa cloud that has one or more physical processors: "When you speak to Alexa, a

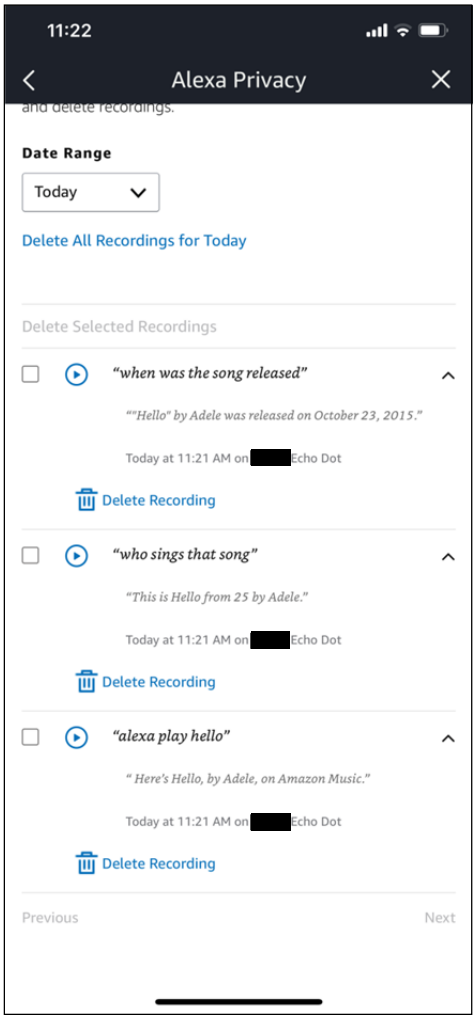
'681 Patent Claim Language	Exemplary Alexa Products
	recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask "Alexa, play top hits on Amazon Music" we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2. Below are more examples. "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles. "The correct hardware configuration and audio processing algorithms can improve your product's listening sensitivity for the wake word Alexa." *Audio Hardware Configurations* at Hardware and Audio Algorithms.
[41.c] accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;	Alexa Products accumulate short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation. 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.

'681 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "Alexa will soon be able to have more natural conversations with users, thanks to something called 'context carryover.' This means that Alexa will be able to understand follow-up questions and respond appropriately..." *Alexa will soon gain a memory*. "Each domain has its own schema to represent slots. The user starts off by asking weather in a city, follows it up with a question about restaurant that serves mexican cuisine, finally she asks about directions to the restaurant...[this] showcase[s] the challenges of multi-domain system...we present a neural network architecture that addresses the challenge..."*Contextual Slot Carryover for Disparate Schemas.*
[41.d] accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;	Alexa Products accumulate long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user. An example is accumulating knowledge about one or more past conversations with the user as shown below in a screenshot of Alexa app's "Review Voice History" page below.

'681 Patent Claim Language	Exemplary Alexa Products
	 As another example, Alexa machine learning scientists "use collaborative filtering...to help resolve ambiguous requests. For instance, if a customer says, "Alexa, play 'Hello,'" his or her music-listening history should indicate whether the song intended is the one by Adele or the one by Lionel Richie." *Interaction Histories.* More examples are below. Alexa uses "interaction history to improve customers' Alexa experiences." *Interaction Histories.* "Conversations are not isolated occurrences. They build on previous questions and answers, whether from five minutes or five days ago. When appropriate, voice-first experiences should account for prior context to make the conversation more engaging or relevant to the user." *What is Conversational AI?*
[41.e] identify a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge;	Alexa Products identify a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge. An example is identifying a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge shown below in a screenshot of Alexa app's "Review Voice History" page below.

'681 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "[C]ontext modeling facilitates a more efficient dialogue by enabling users to reference previous entities and/or intents." *Interaction Histories* (internal quotations omitted). "Capturing and modeling context accurately is critical...even seemingly simple requests...can have multiple valid responses depending on surrounding context." *Interaction Histories* (internal quotations omitted). "The system consists of two parts, Gupta explained: (1) an encoder that produces summary vectors from conversation data, and (2) a decoder that calculates a confidence score for each slot. The two together "learn" slot names and all of their possible values to create embeddings — i.e., points in geometric space that represent strings of words — and then leverage a long short-term memory (LSTM) encoder (a neural network that accounts for the sequencing of data) to identify words helpful in determining candidate slot mapping." *AI helps Alexa*.
[41.f] infer additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in	Alexa Products infer additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context.

'681 Patent Claim Language	Exemplary Alexa Products
response to determining that the utterance contains insufficient information to complete a request in the identified context;	An example is inferring additional information about the utterance shown below in a screenshot of Alexa app's "Review Voice History" page below.  Below are more examples. "The system consists of two parts, Gupta explained: (1) an encoder that produces summary vectors from conversation data, and (2) a decoder that calculates a confidence score for each slot. The two together "learn" slot names and all of their possible values to create embeddings — i.e., points in geometric space that represent strings of words — and then leverage a long short-term memory (LSTM) encoder (a neural network that accounts for the sequencing of data) to identify words helpful in determining candidate slot mapping." *AI helps Alexa*. "[C]ontext modeling facilitates a more efficient dialogue by enabling users to reference previous entities and/or intents." (internal quotations omitted) *Interaction Histories.* Alexa "uses interaction history to improve customers' Alexa experiences." *Interaction Histories.*
[41.g] establish an intended meaning for the utterance within the identified identify a context based	Alexa Products establish an intended meaning for the utterance within the identified identify a context based on the additional information

'681 Patent Claim Language	Exemplary Alexa Products
on the additional information inferred about the utterance; and	inferred about the utterance. An example is establishing an intended meaning for the utterance within the identified context shown below in a screenshot of Alexa app's "Review Voice History" page below.  Below are more examples. "Capturing and modeling context accurately is critical...even seemingly simple requests...can have multiple valid responses depending on surrounding context." (internal quotations omitted) *Interaction Histories.* "Today, Alexa analyzes the semantic content of utterances according to the categories domain, intent, and slot." *How Alexa is Learning*.
[41.h] generate a response to the utterance based on the intended meaning established within the identified context.	Alexa Products generate a response to the utterance based on the intended meaning established within the identified context. An example is generating a response shown below in a screenshot of Alexa app's "Review Voice History" page below.

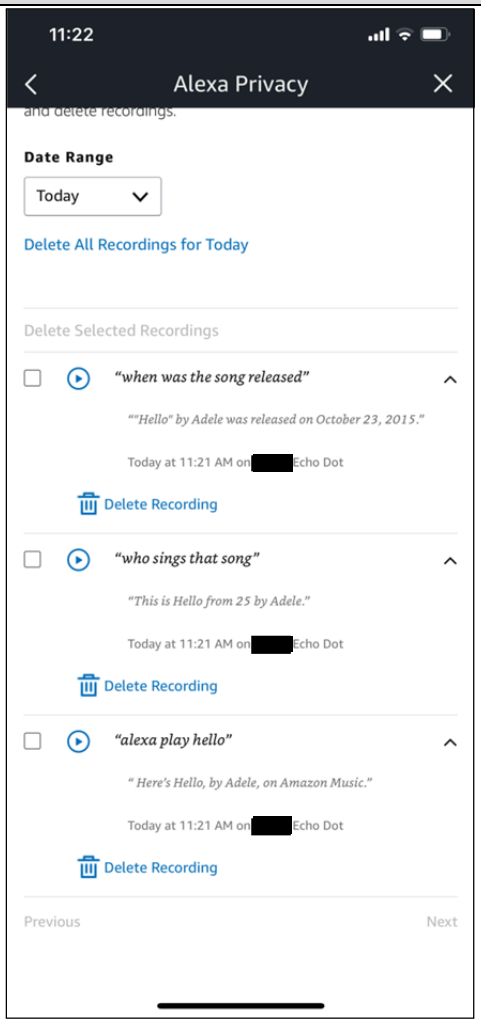
'681 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "We are providing this more natural way of engaging with Alexa by adding deep learning models to our spoken language understanding (SLU) pipeline that allows us to carry customers' intent and entities within and across domains (i.e., between weather and traffic)." *Making Alexa More Friction-Free*. "Alexa will soon be able to have more natural conversations with users, thanks to something called 'context carryover.' This means that Alexa will be able to understand follow-up questions and respond appropriately..." *Alexa will soon gain a memory*.

EXHIBIT E

US009015049B2

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

(10) **Patent No.: US 9,015,049 B2**
(45) **Date of Patent: Apr. 21, 2015**

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

(71) Applicant: **VoiceBox Technologies Corporation**, 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**, Everett, 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 0 days.

(21) Appl. No.: **13/987,645**

(22) Filed: **Aug. 19, 2013**

(65) **Prior Publication Data**

US 2013/0339022 A1 Dec. 19, 2013

Related U.S. Application Data

(60) Division of application No. 13/251,712, filed on Oct. 3, 2011, now Pat. No. 8,515,765, and a continuation of application No. 11/580,926, filed on Oct. 16, 2006, now Pat. No. 8,073,681.

(51) **Int. Cl.**
G10L 15/00 (2013.01)
G10L 25/00 (2013.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G06F 3/167*** (2013.01); ***G10L 15/22*** (2013.01); *G10L 15/183* (2013.01); *G10L 2015/0631* (2013.01)

(58) **Field of Classification Search**
CPC G10L 15/22; G10L 15/18; G10L 15/1815; G10L 15/1822; G10L 15/183; G10L 15/187; G10L 15/19; G10L 15/193; G10L 15/197; G06F 3/167
USPC 704/9, 270–275
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
(Continued)

FOREIGN PATENT DOCUMENTS

EP 1 320 043 A2 6/2003
EP 1 646 037 4/2006
(Continued)

OTHER PUBLICATIONS

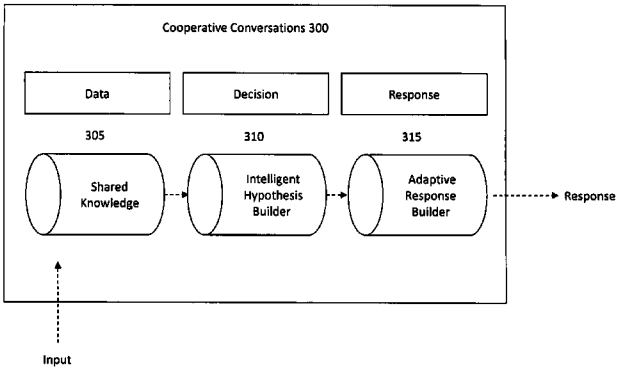
Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., Jul. 28, 2002, 1 page.
(Continued)

Primary Examiner — Eric Yen
(74) *Attorney, Agent, or Firm* — Pillsbury Winthrop Shaw Pittman 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.

20 Claims, 3 Drawing Sheets



(51) **Int. Cl.**

G06F 17/27	(2006.01)
G06F 3/16	(2006.01)
G10L 15/22	(2006.01)
G10L 15/183	(2013.01)
G10L 15/06	(2013.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

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
5,475,733	A		12/1995	Eisdorfer et al. 379/52
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
5,563,937	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
5,617,407	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. 704/270
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
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 et al. 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
5,774,859	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
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
5,832,221	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
5,867,817	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
5,902,347	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 et al. 707/1
5,926,784	A		7/1999	Richardson et al. 704/9
5,933,822	A		8/1999	Braden-Harder et al. 707/5
5,953,393	A		9/1999	Culbreth et al. 379/88.25
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
5,987,404	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
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
6,035,267	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
6,076,059	A		6/2000	Glickman et al. 704/260
6,078,886	A		6/2000	Dragosh et al. 704/270
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
6,122,613	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
6,144,938	A		11/2000	Surace et al. 704/257
6,154,526	A		11/2000	Dahlke et al. 379/88.03
6,160,883	A		12/2000	Jackson et al. 379/230
6,167,377	A		12/2000	Gillick et al. 704/240
6,173,266	B1		1/2001	Marx et al. 704/270
6,173,279	B1		1/2001	Levin et al. 707/5
6,175,858	B1		1/2001	Bulfer et al. 709/206
6,185,535	B1		2/2001	Hedin et al. 704/270
6,188,982	B1		2/2001	Chiang 704/256
6,192,110	B1		2/2001	Abella et al. 379/88.01
6,192,338	B1		2/2001	Haszto et al. 704/257
6,195,634	B1		2/2001	Dudemaine et al. 704/231
6,195,651	B1		2/2001	Handel et al. 707/2
6,199,043	B1		3/2001	Happ 704/272
6,208,964	B1		3/2001	Sabourin 704/244
6,208,972	B1		3/2001	Grant et al. 704/275
6,219,346	B1		4/2001	Maxemchuk 370/338
6,219,643	B1		4/2001	Cohen et al. 704/257
6,226,612	B1		5/2001	Srenger et al. 704/256
6,233,556	B1		5/2001	Teunen et al. 4/250
6,233,559	B1		5/2001	Balakrishnan 704/275
6,233,561	B1		5/2001	Junqua et al. 704/277
6,236,968	B1		5/2001	Kanevsky et al. 704/275
6,246,981	B1	*	6/2001	Papineni et al. 704/235
6,246,990	B1		6/2001	Happ 704/275
6,266,636	B1		7/2001	Kosaka et al. 704/244
6,269,336	B1		7/2001	Ladd et al. 704/270
6,272,455	B1		8/2001	Hoshen et al. 704/1
6,275,231	B1		8/2001	Obradovich 345/349
6,278,377	B1		8/2001	DeLine et al. 340/815.4
6,278,968	B1		8/2001	Franz et al. 704/3
6,288,319	B1		9/2001	Catona 84/609
6,292,767	B1		9/2001	Jackson et al. 704/1
6,301,560	B1		10/2001	Masters 704/251
6,308,151	B1		10/2001	Smith 704/235
6,311,159	B1	*	10/2001	Van Tichelen et al. 704/275
6,314,402	B1		11/2001	Monaco et al. 704/275
6,356,869	B1		3/2002	Chapados et al. 704/275
6,362,748	B1		3/2002	Huang 340/901
6,366,882	B1		4/2002	Bijl et al. 704/235
6,366,886	B1		4/2002	Dragosh et al. 704/270.1
6,374,214	B1		4/2002	Friedland et al. 704/235
6,377,913	B1		4/2002	Coffman et al. 704/8
6,381,535	B1		4/2002	Durocher et al. 701/202
6,385,596	B1		5/2002	Wiser et al. 705/51
6,385,646	B1		5/2002	Brown et al. 709/217
6,393,403	B1		5/2002	Majaniemi 704/275
6,393,428	B1		5/2002	Miller et al. 707/102
6,397,181	B1		5/2002	Li et al. 704/256
6,404,878	B1		6/2002	Jackson et al. 379/221.01
6,405,170	B1		6/2002	Phillips et al. 704/270
6,408,272	B1		6/2002	White et al. 704/270.1
6,411,810	B1		6/2002	Maxemchuk 455/453
6,415,257	B1		7/2002	Junqua et al. 704/275
6,418,210	B1		7/2002	Sayko 379/142.15

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,420,975	B1	7/2002	DeLine et al.	340/815.4
6,429,813	B2	8/2002	Feigen	342/357.13
6,430,285	B1	8/2002	Bauer et al.	379/265.01
6,430,531	B1	8/2002	Polish	704/257
6,434,523	B1	8/2002	Monaco	704/257
6,434,524	B1	8/2002	Weber	704/257
6,434,529	B1	8/2002	Walker et al.	704/275
6,442,522	B1	8/2002	Carberry et al.	704/257
6,446,114	B1	9/2002	Bulfer et al.	709/206
6,453,153	B1	9/2002	Bowker et al.	455/67.4
6,453,292	B2	9/2002	Ramaswamy et al.	704/235
6,456,711	B1	9/2002	Cheung et al.	379/265.09
6,456,974	B1	9/2002	Baker et al.	704/270.1
6,466,654	B1	10/2002	Cooper et al.	379/88.01
6,466,899	B1	10/2002	Yano et al.	704/1
6,470,315	B1	10/2002	Netsch et al.	704/256
6,487,494	B2	11/2002	Odinak et al.	701/202
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
6,505,155	B1	1/2003	Vanbuskirk et al.	704/246
6,510,417	B1	1/2003	Woods et al.	704/275
6,513,006	B2	1/2003	Howard et al.	704/257
6,522,746	B1	2/2003	Marchok et al.	379/406.03
6,523,061	B1	2/2003	Halverson et al.	709/202
6,532,444	B1	3/2003	Weber	704/257
6,539,348	B1	3/2003	Bond et al.	704/9
6,549,629	B2	4/2003	Finn et al.	381/92
6,553,372	B1	4/2003	Brassell et al.	707/5
6,556,970	B1	4/2003	Sasaki et al.	704/257
6,556,973	B1	4/2003	Lewin	704/277
6,560,576	B1	5/2003	Cohen et al.	704/270
6,560,590	B1	5/2003	Shwe et al.	706/55
6,567,778	B1	5/2003	Chao Chang et al.	704/257
6,567,797	B1	5/2003	Schuetze et al.	707/2
6,570,555	B1	5/2003	Prevost et al.	345/156
6,570,964	B1	5/2003	Murveit et al.	379/67.1
6,571,279	B1	5/2003	Herz et al.	709/217
6,574,597	B1	6/2003	Mohri et al.	704/251
6,574,624	B1	6/2003	Johnson et al.	707/5
6,578,022	B1	6/2003	Foulger et al.	706/45
6,581,103	B1	6/2003	Dengler	709/231
6,584,439	B1	6/2003	Geilhufe et al.	704/270
6,587,858	B1	7/2003	Strazza	707/102
6,591,239	B1	7/2003	McCall et al.	704/275
6,594,257	B1	7/2003	Doshi et al.	370/352
6,594,367	B1	7/2003	Marash et al.	381/92
6,598,018	B1	7/2003	Junqua	704/251
6,601,026	B2	7/2003	Appelt et al.	704/9
6,604,075	B1	8/2003	Brown et al.	704/270.1
6,604,077	B2	8/2003	Dragosh et al.	704/270.1
6,606,598	B1	8/2003	Holthouse et al.	704/275
6,611,692	B2	8/2003	Raffel et al.	455/552
6,614,773	B1	9/2003	Maxemchuk	370/337
6,615,172	B1	9/2003	Bennett et al.	704/257
6,622,119	B1	9/2003	Ramaswamy et al.	704/9
6,629,066	B1	9/2003	Jackson et al.	704/9
6,631,346	B1	10/2003	Karaorman et al.	704/9
6,631,351	B1	10/2003	Ramachandran et al.	704/270
6,633,846	B1	10/2003	Bennett et al.	704/257
6,636,790	B1	10/2003	Lightner et al.	701/33
6,643,620	B1	11/2003	Contolini et al.	704/270
6,647,363	B2 *	11/2003	Claassen	704/1
6,650,747	B1	11/2003	Bala et al.	379/265.06
6,658,388	B1 *	12/2003	Kleindienst et al.	704/275
6,678,680	B1	1/2004	Woo	707/6
6,681,206	B1	1/2004	Gorin et al.	704/243
6,691,151	B1	2/2004	Cheyer 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
6,708,150	B1	3/2004	Hirayama et al.	704/243
6,721,001	B1	4/2004	Berstis	348/231.3
6,721,706	B1	4/2004	Strubbe et al.	704/275
6,735,592	B1	5/2004	Neumann et al.	707/101
6,741,931	B1	5/2004	Kohut et al.	701/209
6,742,021	B1	5/2004	Halverson et al.	709/218
6,745,161	B1	6/2004	Arnold et al.	704/7
6,751,591	B1	6/2004	Gorin et al.	704/257
6,751,612	B1	6/2004	Schuetze et al.	707/4
6,754,485	B1	6/2004	Obradovich et al.	455/414.1
6,754,627	B2	6/2004	Woodward	704/235
6,757,544	B2	6/2004	Rangarajan et al.	455/456.1
6,757,718	B1	6/2004	Halverson et al.	709/218
6,795,808	B1	9/2004	Strubbe et al.	704/275
6,801,604	B2	10/2004	Maes et al.	379/88.17
6,801,893	B1	10/2004	Backfried et al.	704/257
6,829,603	B1	12/2004	Wolf et al.	707/5
6,832,230	B1	12/2004	Zilliacus et al.	707/203
6,833,848	B1	12/2004	Wolff et al.	345/719
6,850,603	B1	2/2005	Eberle et al.	379/88.16
6,856,990	B2	2/2005	Barile et al.	707/10
6,865,481	B2	3/2005	Kawazoe et al.	701/211
6,868,380	B2	3/2005	Kroeker	704/240
6,877,001	B2	4/2005	Wolf et al.	707/3
6,877,134	B1	4/2005	Fuller et al.	715/500.1
6,901,366	B1	5/2005	Kuhn et al.	704/275
6,910,003	B1	6/2005	Arnold et al.	704/4
6,912,498	B2	6/2005	Stevens et al.	704/235
6,934,756	B2	8/2005	Maes	709/227
6,937,977	B2	8/2005	Gerson	704/201
6,941,266	B1 *	9/2005	Gorin et al.	704/257
6,944,594	B2	9/2005	Busayapongchai et al.	704/275
6,950,821	B2	9/2005	Faybishenko et al.	707/10
6,954,755	B2	10/2005	Reisman	707/10
6,959,276	B2	10/2005	Droppo et al.	704/226
6,968,311	B2	11/2005	Knockeart et al.	704/270
6,973,387	B2	12/2005	Masclet et al.	701/211
6,975,993	B1	12/2005	Keiller	704/275
6,980,092	B2	12/2005	Turnbull et al.	340/425.5
6,983,055	B2	1/2006	Luo	381/313
6,990,513	B2	1/2006	Belfiore et al.	709/203
6,996,531	B2	2/2006	Korall et al.	704/270
7,003,463	B1	2/2006	Maes et al.	704/270.1
7,016,849	B2	3/2006	Arnold et al.	704/275
7,020,609	B2	3/2006	Thrift et al.	704/270.1
7,024,364	B2	4/2006	Guerra et al.	704/270
7,027,586	B2	4/2006	Bushey et al.	379/265.09
7,027,975	B1	4/2006	Pazandak et al.	704/9
7,035,415	B2	4/2006	Belt et al.	381/92
7,036,128	B1	4/2006	Julia et al.	719/317
7,043,425	B2	5/2006	Pao	704/211
7,054,817	B2	5/2006	Shao	704/270
7,058,890	B2	6/2006	George et al.	715/728
7,062,488	B1	6/2006	Reisman	707/8
7,069,220	B2	6/2006	Coffman et al.	704/275
7,072,834	B2	7/2006	Zhou	704/244
7,082,469	B2	7/2006	Gold et al.	709/231
7,092,928	B1	8/2006	Elad et al.	706/60
7,107,210	B2	9/2006	Deng et al.	704/226
7,107,218	B1	9/2006	Preston	704/270
7,110,951	B1	9/2006	Lemelson et al.	704/270
7,127,395	B1 *	10/2006	Gorin et al.	704/257
7,127,400	B2	10/2006	Koch	704/270.1
7,130,390	B2	10/2006	Abburi	379/88.17
7,136,875	B2	11/2006	Anderson et al.	707/104.1
7,137,126	B1	11/2006	Coffman et al.	719/328
7,143,037	B1	11/2006	Chestnut	704/251
7,143,039	B1	11/2006	Stifelman et al.	704/270
7,146,319	B2	12/2006	Hunt	704/254
7,165,028	B2	1/2007	Gong	704/233
7,177,798	B2	2/2007	Hsu et al.	704/9
7,184,957	B2	2/2007	Brookes et al.	704/246
7,190,770	B2	3/2007	Ando et al.	379/88.01
7,197,069	B2	3/2007	Agazzi et al.	375/233
7,197,460	B1	3/2007	Gupta et al.	704/270.1
7,203,644	B2	4/2007	Anderson et al.	704/246
7,206,418	B2	4/2007	Yang et al.	381/92
7,207,011	B2	4/2007	Mulvey et al.	715/812
7,228,276	B2	6/2007	Omote et al.	704/243
7,231,343	B1	6/2007	Treadgold et al.	704/9
7,236,923	B1	6/2007	Gupta	704/9

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,272,212	B2	9/2007	Eberle et al.	379/88.17
7,277,854	B2	10/2007	Bennett et al.	704/257
7,283,829	B2	10/2007	Christenson et al.	455/461
7,283,951	B2	10/2007	Marchisio et al.	704/9
7,289,606	B2	10/2007	Sibal et al.	379/52
7,299,186	B2	11/2007	Kuzunuki et al.	704/270.1
7,301,093	B2	11/2007	Sater et al.	84/615
7,305,381	B1	12/2007	Poppink et al.	1/1
7,321,850	B2	1/2008	Wakita	704/10
7,328,155	B2	2/2008	Endo et al.	704/251
7,337,116	B2	2/2008	Charlesworth et al.	704/254
7,340,040	B1	3/2008	Saylor et al.	379/67.1
7,366,285	B2	4/2008	Parolkar et al.	379/88.17
7,366,669	B2	4/2008	Nishitani et al.	704/256
7,376,645	B2	5/2008	Bernard	707/3
7,386,443	B1	6/2008	Parthasarathy et al.	704/201
7,398,209	B2	7/2008	Kennewick et al.	704/255
7,406,421	B2	7/2008	Odinak et al.	704/275
7,415,414	B2	8/2008	Azara et al.	704/270
7,421,393	B1	9/2008	Di Fabbrizio et al.	704/275
7,424,431	B2	9/2008	Greene et al.	704/270
7,447,635	B1	11/2008	Konopka et al.	704/275
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
7,472,075	B2	12/2008	Odinak et al.	705/26
7,478,036	B2	1/2009	Shen et al.	704/9
7,487,088	B1	2/2009	Gorin et al.	704/240
7,487,110	B2	2/2009	Bennett et al.	705/26
7,493,259	B2	2/2009	Jones et al.	704/257
7,493,559	B1	2/2009	Wolff et al.	715/727
7,502,672	B1	3/2009	Kolls	701/29
7,502,738	B2	3/2009	Kennewick et al.	704/257
7,516,076	B2	4/2009	Walker et al.	704/275
7,529,675	B2	5/2009	Maes	704/270.1
7,536,297	B2	5/2009	Byrd et al.	704/10
7,536,374	B2	5/2009	Au	706/55
7,542,894	B2	6/2009	Murata	704/9
7,546,382	B2	6/2009	Healey et al.	709/246
7,558,730	B2	7/2009	Davis et al.	704/235
7,574,362	B2	8/2009	Walker et al.	704/275
7,577,244	B2	8/2009	Taschereau	379/218.01
7,606,708	B2	10/2009	Hwang	704/257
7,620,549	B2	11/2009	Di Cristo et al.	704/257
7,634,409	B2	12/2009	Kennewick et al.	704/257
7,640,006	B2	12/2009	Portman et al.	455/412.1
7,640,160	B2	12/2009	Di Cristo et al.	704/257
7,640,272	B2	12/2009	Mahajan et al.	707/104.1
7,676,365	B2	3/2010	Hwang et al.	704/240
7,676,369	B2	3/2010	Fujimoto et al.	704/270
7,684,977	B2	3/2010	Morikawa	704/211
7,693,720	B2	4/2010	Kennewick et al.	704/275
7,729,916	B2	6/2010	Coffman et al.	704/270
7,729,918	B2	6/2010	Walker et al.	704/275
7,729,920	B2	6/2010	Chaar et al.	704/275
7,734,287	B2	6/2010	Ying	455/423
7,748,021	B2	6/2010	Obradovich	725/105
7,788,084	B2	8/2010	Brun et al.	704/7
7,801,731	B2	9/2010	Odinak et al.	704/275
7,809,570	B2	10/2010	Kennewick et al.	704/257
7,818,176	B2	10/2010	Freeman et al.	704/270
7,831,426	B2	11/2010	Bennett	704/252
7,831,433	B1	11/2010	Belvin et al.	704/275
7,873,519	B2	1/2011	Bennett	704/257
7,873,523	B2	1/2011	Potter et al.	704/275
7,881,936	B2	2/2011	Longe et al.	704/257
7,890,324	B2	2/2011	Bangalore et al.	704/231
7,894,849	B2	2/2011	Kass et al.	455/550.1
7,902,969	B2	3/2011	Obradovich	340/439
7,917,367	B2	3/2011	Di Cristo et al.	704/270.1
7,920,682	B2	4/2011	Byrne et al.	379/88.18
7,949,529	B2	5/2011	Weider et al.	704/270
7,949,537	B2	5/2011	Walker et al.	704/275
7,953,732	B2	5/2011	Frank et al.	707/724
7,974,875	B1	7/2011	Quilici et al.	705/14.4
7,983,917	B2	7/2011	Kennewick et al.	704/257
7,984,287	B2	7/2011	Gopalakrishnan et al.	713/100
8,005,683	B2	8/2011	Tessel et al.	704/275
8,015,006	B2	9/2011	Kennewick et al.	704/236
8,060,367	B2	11/2011	Keaveney	704/247
8,069,046	B2	11/2011	Kennewick et al.	704/257
8,073,681	B2	12/2011	Baldwin et al.	704/9
8,077,975	B2	12/2011	Ma et al.	382/187
8,082,153	B2	12/2011	Coffman et al.	704/270
8,086,463	B2	12/2011	Ativanichayaphong et al.	704/275
8,112,275	B2	2/2012	Kennewick et al.	704/240
8,140,327	B2	3/2012	Kennewick et al.	704/226
8,140,335	B2	3/2012	Kennewick et al.	704/257
8,145,489	B2	3/2012	Freeman et al.	704/257
8,150,694	B2	4/2012	Kennewick et al.	704/257
8,155,962	B2	4/2012	Kennewick et al.	704/257
8,170,867	B2	5/2012	Germain	704/9
8,195,468	B2	6/2012	Weider et al.	704/275
8,219,399	B2	7/2012	Lutz et al.	704/270
8,219,599	B2	7/2012	Tunstall-Pedoe	707/955
8,224,652	B2	7/2012	Wang et al.	704/275
8,255,224	B2	8/2012	Singleton et al.	704/275
8,326,627	B2	12/2012	Kennewick et al.	704/257
8,326,634	B2	12/2012	Di Cristo et al.	704/270.1
8,326,637	B2	12/2012	Baldwin et al.	704/275
8,332,224	B2	12/2012	Di Cristo et al.	704/257
8,370,147	B2	2/2013	Kennewick et al.	704/257
8,447,607	B2	5/2013	Weider et al.	704/250
8,452,598	B2	5/2013	Kennewick et al.	704/257
8,719,026	B2	5/2014	Kennewick et al.	704/257
8,731,929	B2	5/2014	Kennewick et al.	704/257
8,738,380	B2	5/2014	Baldwin et al.	704/257
8,849,652	B2	9/2014	Weider et al.	704/9
8,849,670	B2	9/2014	Di Cristo et al.	704/270.1
8,886,536	B2	11/2014	Freeman et al.	704/257
2001/0041980	A1	11/2001	Howard et al.	704/270
2001/0049601	A1	12/2001	Kroeker et al.	704/254
2001/0054087	A1	12/2001	Flom et al.	709/218
2002/0015500	A1	2/2002	Belt et al.	381/66
2002/0022927	A1	2/2002	Lemelson et al.	701/301
2002/0029261	A1	3/2002	Shibata	709/219
2002/0032752	A1	3/2002	Gold et al.	709/218
2002/0035501	A1	3/2002	Handel et al.	705/10
2002/0040297	A1	4/2002	Tsiao et al.	704/257
2002/0049535	A1	4/2002	Rigo et al.	701/211
2002/0049805	A1	4/2002	Yamada et al.	709/202
2002/0059068	A1*	5/2002	Rose et al.	704/246
2002/0065568	A1	5/2002	Silfvast et al.	700/94
2002/0067839	A1	6/2002	Heinrich	381/110
2002/0069059	A1	6/2002	Smith	704/257
2002/0069071	A1	6/2002	Knockeart et al.	704/275
2002/0082911	A1	6/2002	Dunn et al.	705/14
2002/0087312	A1*	7/2002	Lee et al.	704/251
2002/0087326	A1	7/2002	Lee et al.	704/270.1
2002/0087525	A1	7/2002	Abbott et al.	707/3
2002/0107694	A1	8/2002	Lerg	704/273
2002/0120609	A1	8/2002	Lang et al.	707/1
2002/0124050	A1	9/2002	Middeljans	709/203
2002/0133354	A1*	9/2002	Ross et al.	704/275
2002/0133402	A1	9/2002	Faber et al.	705/14
2002/0135618	A1	9/2002	Maes et al.	345/767
2002/0138248	A1	9/2002	Corston-Oliver et al.	704/1
2002/0143532	A1	10/2002	McLean et al.	704/235
2002/0143535	A1	10/2002	Kist et al.	704/251
2002/0161646	A1	10/2002	Gailey et al.	705/14
2002/0173333	A1	11/2002	Buchholz et al.	455/527
2002/0173961	A1	11/2002	Guerra	704/258
2002/0184373	A1	12/2002	Maes	709/228
2002/0188602	A1	12/2002	Stubler et al.	707/3
2002/0198714	A1	12/2002	Zhou	704/252
2003/0014261	A1	1/2003	Kageyama	704/275
2003/0016835	A1	1/2003	Elko et al.	381/92
2003/0046346	A1	3/2003	Mumick et al.	709/205
2003/0064709	A1	4/2003	Gailey et al.	455/412
2003/0065427	A1	4/2003	Funk et al.	701/1
2003/0069734	A1	4/2003	Everhart	704/275
2003/0088421	A1	5/2003	Maes et al.	704/270.1

(56) **References Cited**

U.S. PATENT DOCUMENTS

2003/0097249 A1 5/2003 Walker et al. 704/1
2003/0110037 A1 6/2003 Walker et al. 704/257
2003/0112267 A1 6/2003 Belrose 345/728
2003/0115062 A1 6/2003 Walker et al. 704/258
2003/0120493 A1 6/2003 Gupta 704/270.1
2003/0135488 A1 7/2003 Amir et al. 707/3
2003/0144846 A1 7/2003 Denenberg et al. 704/277
2003/0158731 A1 8/2003 Falcon et al. 704/231
2003/0161448 A1 8/2003 Parolkar et al. 379/88.17
2003/0182132 A1 9/2003 Niemoeller 704/275
2003/0204492 A1 10/2003 Wolf et al. 707/3
2003/0206640 A1 11/2003 Malvar et al. 381/93
2003/0212550 A1 11/2003 Ubale 704/215
2003/0212558 A1 11/2003 Matula 704/260
2003/0212562 A1 11/2003 Patel et al. 704/275
2003/0225825 A1 12/2003 Healey et al. 709/203
2003/0236664 A1 12/2003 Sharma 704/251
2004/0006475 A1 1/2004 Ehlen et al. 704/270.1
2004/0010358 A1 1/2004 Oesterling et al. 701/49
2004/0025115 A1 2/2004 Sienel et al. 715/513
2004/0044516 A1 3/2004 Kennewick et al. 704/5
2004/0098245 A1 5/2004 Walker et al. 704/1
2004/0117179 A1 6/2004 Balasuriya 704/231
2004/0117804 A1 6/2004 Scahill et al. 719/320
2004/0122674 A1 6/2004 Bangalore et al. 704/276
2004/0140989 A1 7/2004 Papageorge 345/700
2004/0158555 A1 8/2004 Seedman et al. 707/3
2004/0166832 A1 8/2004 Portman et al. 455/412.1
2004/0167771 A1 8/2004 Duan et al. 704/10
2004/0172258 A1 9/2004 Dominach et al. 704/277
2004/0193408 A1 9/2004 Hunt 704/209
2004/0193420 A1 9/2004 Kennewick et al. 704/257
2004/0199375 A1 10/2004 Ehsani et al. 704/4
2004/0205671 A1 10/2004 Sukehiro et al. 715/532
2004/0243417 A9 12/2004 Pitts, III et al. 704/276
2005/0015256 A1 1/2005 Kargman 704/272
2005/0021331 A1* 1/2005 Huang et al. 704/231
2005/0021334 A1 1/2005 Iwahashi 704/240
2005/0021470 A1 1/2005 Martin et al. 705/51
2005/0021826 A1 1/2005 Kumar 709/232
2005/0033574 A1 2/2005 Kim et al. 704/251
2005/0033582 A1 2/2005 Gadd et al. 704/277
2005/0043940 A1 2/2005 Elder 704/9
2005/0114116 A1 5/2005 Fiedler 704/201
2005/0125232 A1 6/2005 Gadd 704/270.1
2005/0131673 A1 6/2005 Koizumi et al. 704/2
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
2005/0144013 A1* 6/2005 Fujimoto et al. 704/277
2005/0216254 A1 9/2005 Gupta et al. 704/9
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
2006/0074671 A1* 4/2006 Farmaner et al. 704/257
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
2007/0038445 A1* 2/2007 Helbing et al. 704/235
2007/0043569 A1 2/2007 Potter, III et al. 704/270
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
2007/0061067 A1 3/2007 Zeinstra et al. 701/200
2007/0061735 A1 3/2007 Hoffberg et al. 715/744
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
2007/0094003 A1* 4/2007 Huang et al. 704/3
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
2007/0203736 A1 8/2007 Ashton 705/1
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
2008/0147410 A1 6/2008 Odinak 704/270.1
2008/0154604 A1 6/2008 Sathish et al. 704/257
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
2009/0307031 A1 12/2009 Winkler et al. 705/7
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
2010/0145700 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
2011/0112827 A1 5/2011 Kennewick et al. 704/9
2011/0112921 A1 5/2011 Kennewick et al. 705/26.1
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
2013/0054228 A1 2/2013 Baldwin et al. 704/9
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
2013/0297293 A1 11/2013 Di Cristo at al. 704/9
2013/0304473 A1 11/2013 Baldwin et al. 704/257
2014/0012577 A1 1/2014 Freeman at al. 704/249
2014/0108013 A1 4/2014 Di Cristo et al. 704/254
2014/0156278 A1 6/2014 Kennewick et al. 704/254

(56) **References Cited**

U.S. PATENT DOCUMENTS

2014/0249821	A1	9/2014	Kennewick et al.	704/257
2014/0249822	A1	9/2014	Baldwin et al.	704/257
2014/0288934	A1	9/2014	Kennewick et al.	704/254

FOREIGN PATENT DOCUMENTS

JP	2006-146881	6/2006
JP	2008-027454	2/2008
JP	2008-139928	6/2008
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.

Cheyer 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.

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

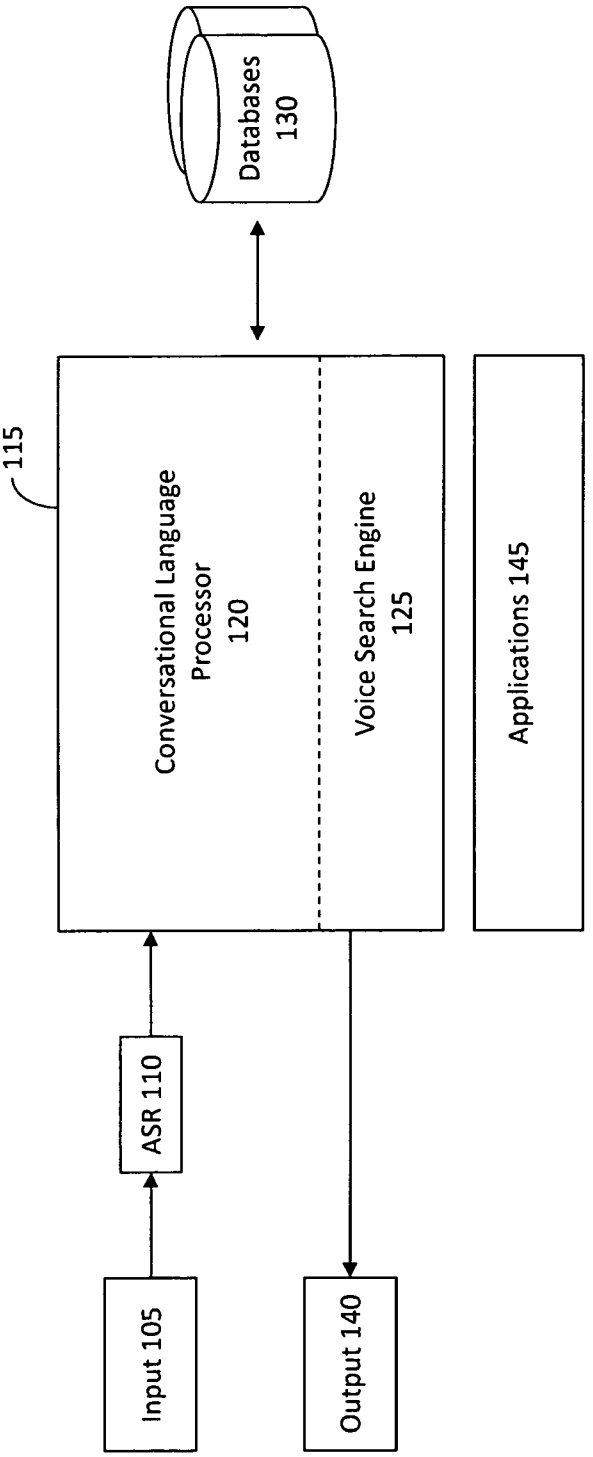

Databases
130
115
Conversational Language
Processor
120
Voice Search Engine
125
Applications 145
ASR 110
Input 105
Output 140

Figure 1

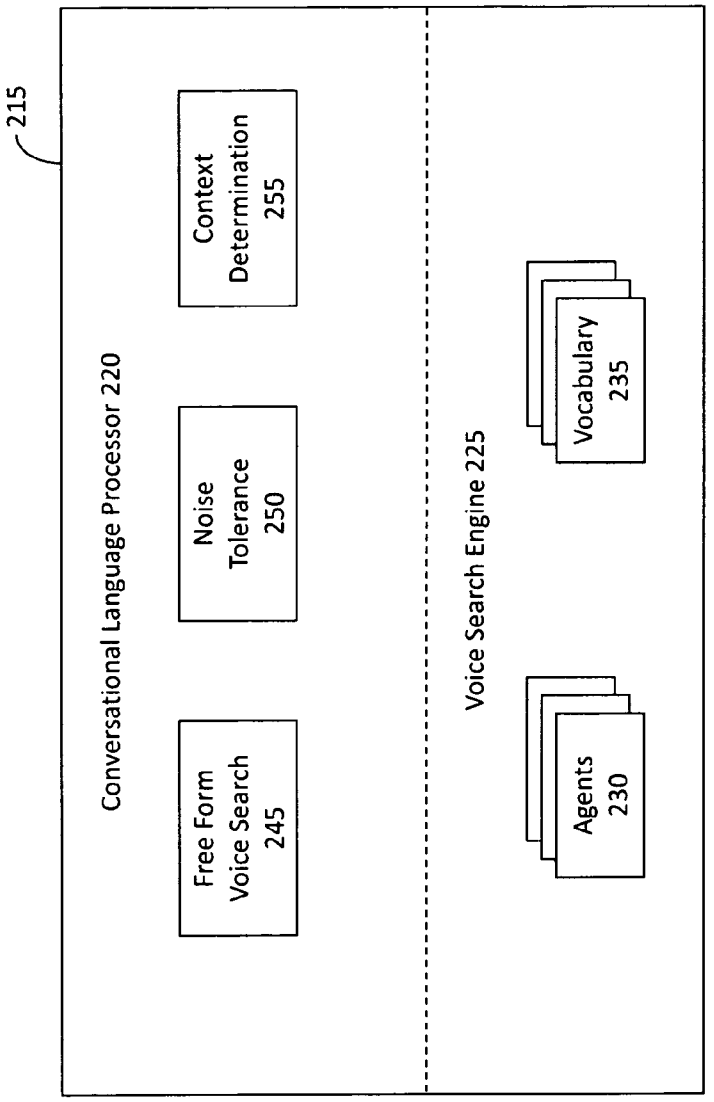

215
Conversational Language Processor 220
Free Form Voice Search 245
Noise Tolerance 250
Context Determination 255
Voice Search Engine 225
Agents 230
Vocabulary 235

Figure 2

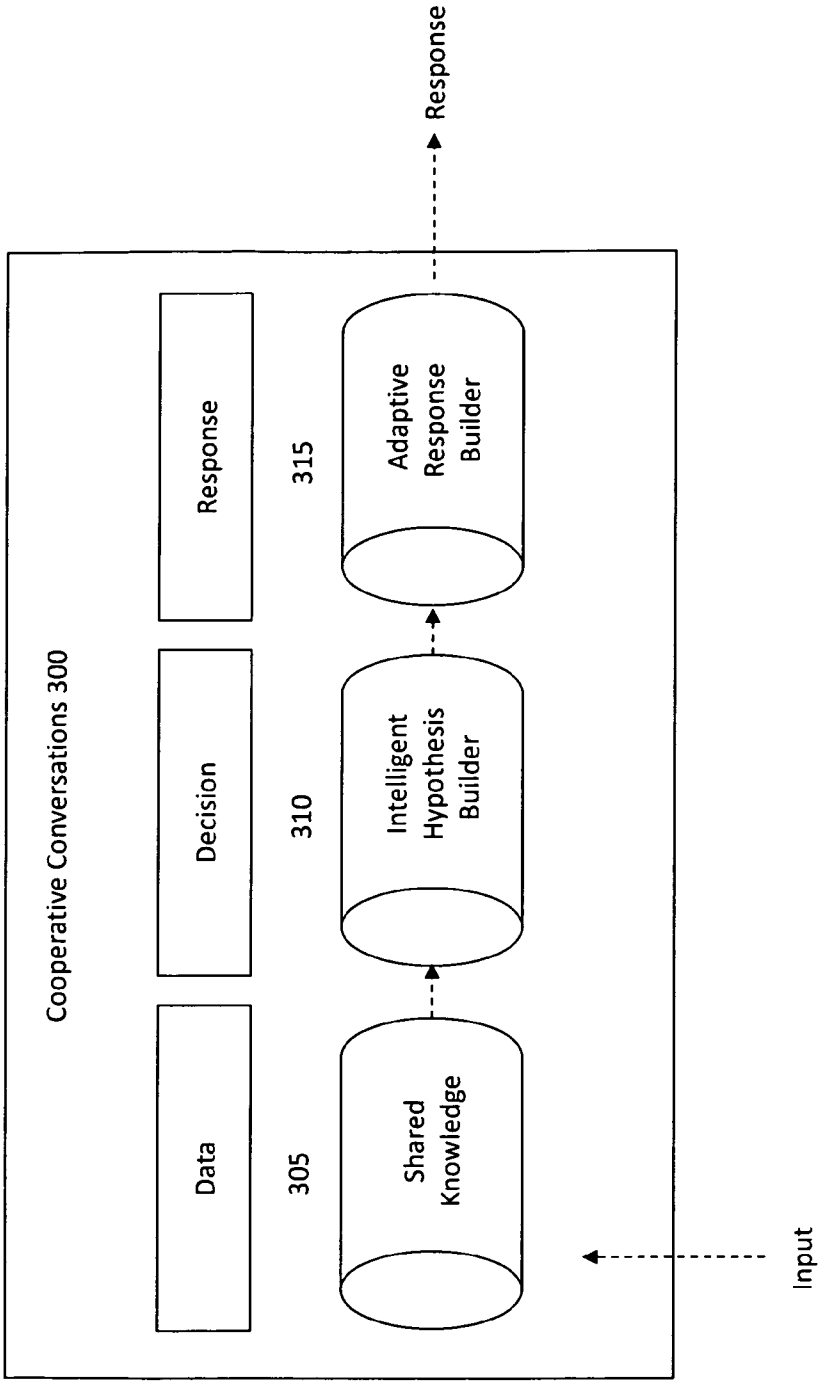

Response
Cooperative Conversations 300
Response
Decision
Data
315
310
305
Adaptive Response Builder
Intelligent Hypothesis Builder
Shared Knowledge
Input

Figure 3

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

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

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). 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, urn, 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 iden-

tify "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 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 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.

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, Oreg. based on the Weather and an environmental proximity between Portland, Oreg. 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, Oreg.?").

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

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 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, 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	Less	More
ALLOCATION		

DIDACTIC

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

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	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 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 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?

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	In Dallas, it's sunny and 90 degrees.

-continued

Interface	
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 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 facilitating conversation-based responses, 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, at the computer system, a natural language utterance during a conversation between a user and the computer system;

identifying, by the computer system, a first model that includes short-term knowledge about the conversation, wherein the short-term knowledge is based on one or more prior natural language utterances received during the conversation;

identifying, by the computer system, based on the short-term knowledge, context information for the natural language utterance;

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

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

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

3. The method of claim **1**, further comprising:

updating, by the computer system, based on information about the natural language utterance, during the conversation, the short-term knowledge in the first model,

wherein the updated short-term knowledge is used to determine subsequent context information for one or more subsequent natural language utterances received during the conversation.

4. The method of claim **1**, further comprising:

identifying, by the computer system, a second model that includes long-term knowledge about one or more prior conversations between the user and the computer system,

wherein determining the context information comprises determining, based on the short-term knowledge and the long-term knowledge, the context information.

5. The method of claim **4**, further comprising:

updating, by the computer system, based on information about the natural language utterance, the long-term knowledge in the second model,

wherein the updated long-term knowledge is used to determine subsequent context information for one or more subsequent natural language utterances received during the conversation.

6. The method of claim **5**, wherein the updated long-term knowledge is used to determine subsequent context information for one or more subsequent conversations.

7. The method of claim **1**, further comprising:

identifying, by the computer system, based on the short-term knowledge, a manner in which the natural language utterance is spoken,

wherein generating the response comprises generating, based on the identified manner and the interpretation of the natural language utterance, the response.

8. The method of claim **7**, further comprising:

identifying, by the computer system, a second model that includes long-term knowledge about one or more prior conversations between the user and the computer system,

wherein identifying the manner comprises identifying, based on the short-term knowledge and the long-term knowledge, the manner.

9. The method of claim 7, wherein generating the response comprises generating, based on a response format associated with the identified manner, the response.

10. The method of claim 1, wherein the natural language utterance and the one or more prior natural language utterances are associated with the user.

11. A system for facilitating conversation-based responses, the system comprising:

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

receive a natural language utterance during a conversation between a user and the system;

identify a first model that includes short-term knowledge about the conversation, wherein the short-term knowledge is based on one or more prior natural language utterances received during the conversation;

identify, based on the short-term knowledge, context information for the natural language utterance;

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

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

12. The system of claim 11, wherein determining the interpretation of the natural language utterance comprises determining, based on the context information, an interpretation of one or more recognized words of the natural language utterance.

13. The system of claim 11, wherein the one or more physical processors are caused to:

update, based on information about the natural language utterance, during the conversation, the short-term knowledge in the first model,

wherein the updated short-term knowledge is used to determine subsequent context information for one or more subsequent natural language utterances received during the conversation.

14. The system of claim 11, wherein the one or more physical processors are caused to:

identify a second model that includes long-term knowledge about one or more prior conversations between the user and the system,

wherein determining the context information comprises determining, based on the short-term knowledge and the long-term knowledge, the context information.

15. The system of claim 14, wherein the one or more physical processors are caused to:

update, based on information about the natural language utterance, the long-term knowledge in the second model,

wherein the updated long-term knowledge is used to determine subsequent context information for one or more subsequent natural language utterances received during the conversation.

16. The system of claim 15, wherein the updated long-term knowledge is used to determine subsequent context information for one or more subsequent conversations.

17. The system of claim 11, wherein the one or more physical processors are caused to:

identify, based on the short-term knowledge, a manner in which the natural language utterance is spoken,

wherein generating the response comprises generating, based on the identified manner and the interpretation of the natural language utterance, the response.

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

identify a second model that includes long-term knowledge about one or more prior conversations between the user and the computer system,

wherein identifying the manner comprises identifying, based on the short-term knowledge and the long-term knowledge, the manner.

19. The system of claim 17, wherein generating the response comprises generating, based on a response format associated with the identified manner, the response.

20. The system of claim 11, wherein the natural language utterance and the one or more prior natural language utterances are associated with the user.

* * * * *

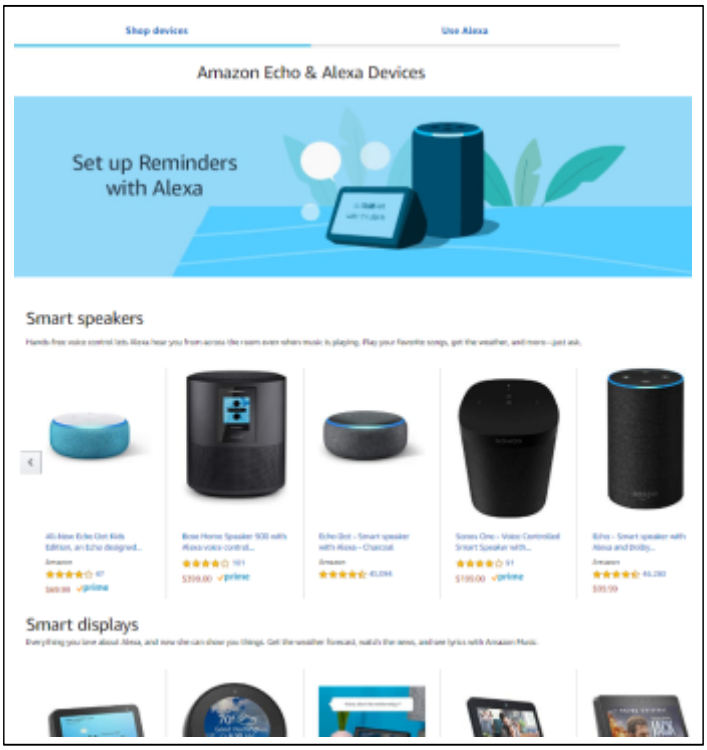
EXHIBIT F

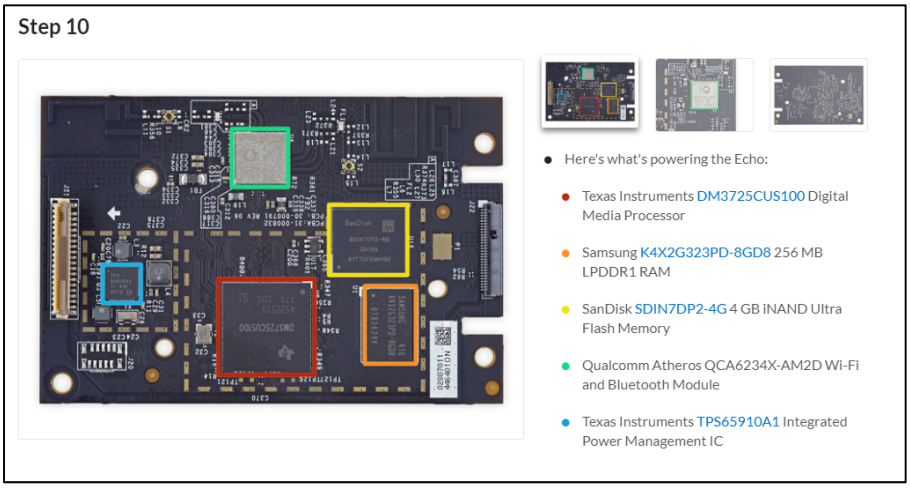
U.S. Pat. No. 9,015,049 ("the '049 patent")
Infringement Chart for Claim 11

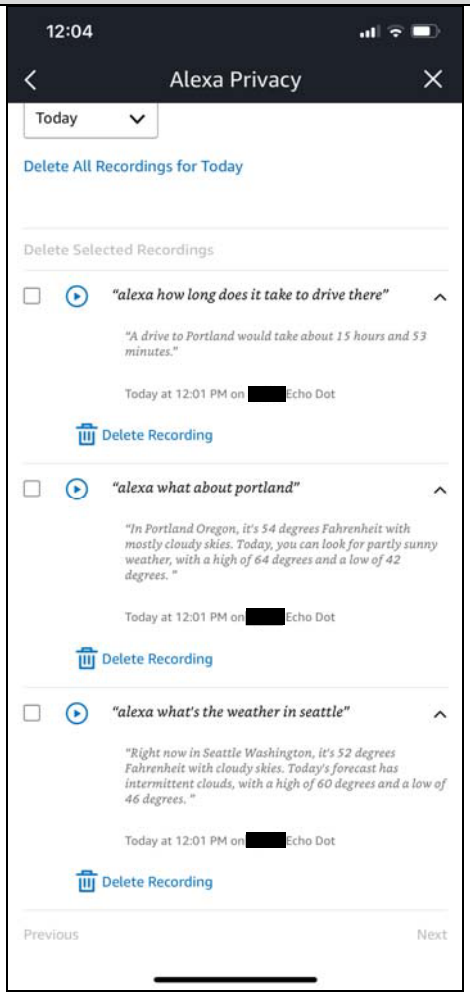
The following chart provides exemplary infringement allegations with respect to an exemplary claim of the '049 patent based on Plaintiff's investigation to date. Plaintiff reserves the right to provide modified and/or more detailed allegations (including with additional claims, products, and evidence) in accordance with its discovery obligations in view of any applicable scheduling order. This chart uses "Alexa Products" to refer to the full scope of instrumentalities identified in the complaint. Any discussion of specific Alexa Products in the following chart is exemplary.

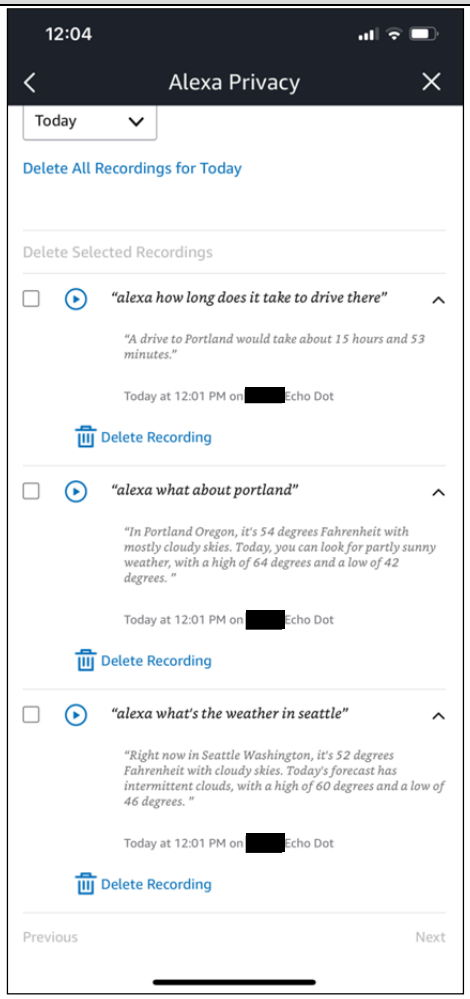
Exemplary information relevant to infringement:

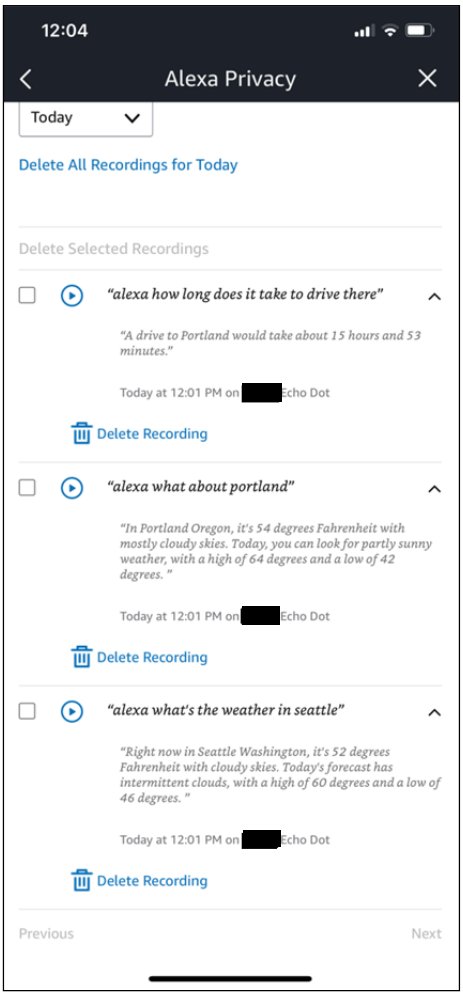
1. *Amazon Echo & Alexa Devices* - Amazon Official Site, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/?ie=UTF8&node=9818047011&ref_=sv_devicesubnav_1 (last visited Jul. 22, 2019)
2. *Alexa and Alexa Device FAQs*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (last visited Jul. 22, 2019)
3. *Use Alexa > Alexa Features > Entertainment*, Amazon.com, https://www.amazon.com/b/ref=aeg_flp_ent_text/ref=s9_acss_bw_cg_aegflp_md1_w?node=17934676011&pf_rd_m=ATVPDKIKX0DER&pf_rd_s=merchandised-search-6&pf_rd_r=52PC5VKKVN8N50VH8CXF&pf_rd_t=101&pf_rd_p=6414bf70-6ba0-4f3e-9138-998f49588b76&pf_rd_i=17934672011 (last visited Jul. 22, 2019)
4. Sarah Perez, *Alexa will soon gain a memory, converse more naturally, and automatically launch skills*, TechCrunch, Apr. 26, 2018, https://techcrunch.com/2018/04/26/alexa-will-soon-gain-a-memory-converse-more-naturally-and-automatically-launch-skills/ ("*Alexa will soon gain a memory*")
5. Ruhi Sarikaya, *Making Alexa More Friction-Free*, Apr. 25, 2018 post on Alexa Blogs, https://developer.amazon.com/blogs/alexa/post/60e1f011-3236-4162-b0f6-509205d354ca/making-alexa-more-friction-free
6. Arpit Gupta, *How Alexa Is Learning to Converse More Naturally*, Aug. 27, 2018 post on Alexa Blogs, https://developer.amazon.com/blogs/alexa/post/15bf7d2a-5e5c-4d43-90ae-c2596c9cc3a6/how-alexa-is-learning-to-converse-more-naturally ("*How Alexa is Learning*")
7. Kyle Wiggers, *Amazon researchers share how AI helps Alexa answer follow-up questions*, VentureBeat, August 27, 2018, https://venturebeat.com/2018/08/27/amazon-researchers-share-how-ai-helps-alexa-answer-follow-up-questions/ ("*AI helps Alexa*")
8. *What is Conversational AI?*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/alexa-skills-kit/conversational-ai (last visited Jul. 22, 2019)
9. *Audio Hardware Configurations*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/docs/alexa-voice-service/audio-hardware-configurations.html (last visited Jul. 22, 2019)
10. *Designing for Conversation – Build for dialog*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/ask-resources/guided/conversational-design-workshop#/lessons/gDcQVF9nOFOtNljrELzARIGN4HuD7LNh?_k=q4qp1a (last visited Jul. 22, 2019)
11. *Amazon Echo Teardown*, IFIXIT, Dec. 16, 2014, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953

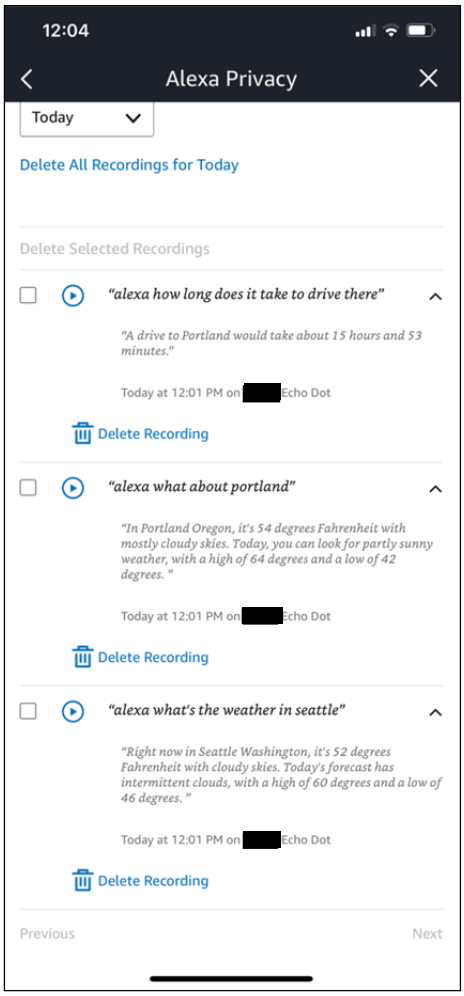
'049 Patent Claim Language	**Exemplary Alexa Products**
11.[pr] A system for facilitating conversation-based responses, the system comprising:	Alexa Products include a system for facilitating conversation-based responses. For example, Alexa Products (e.g. Echo, Echo Dot, Alexa apps, and Alexa cloud) include Alexa.  E.g. *Amazon Echo & Alexa Devices*. "With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?* Alexa is connected to Amazon's Alexa cloud. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1. "When it comes to connecting you to everything you're into, Alexa is at home anywhere. Whether you want to listen to music, books, radio stations, or podcasts, all you have to do is ask. In the mood for games, movies, or a good laugh? Let Alexa connect the dots between all the things you love." *Use Alexa > Alexa Features > Entertainment.*
[11.a] one or more physical processors programmed with one or more computer program instructions such that, when executed, the one or more computer program instructions	Alexa Products include one or more physical processors programmed with one or more computer program instructions such that, when executed, the one or more computer program instructions cause the one or more physical processors to perform the operations recited in the claim.

'049 Patent Claim Language	Exemplary Alexa Products
cause the one or more physical processors to:	For example, the Echo has the physical processors shown below:  E.g. *Amazon Echo Teardown* at Step 10. As another example, Alexa Products include Amazon's Alexa cloud that has one or more physical processors: "When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask "Alexa, play top hits on Amazon Music" we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2. Below are more examples. "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles. "The correct hardware configuration and audio processing algorithms can improve your product's listening sensitivity for the wake word Alexa." *Audio Hardware Configurations* at Hardware and Audio Algorithms.
[11.b] receive a natural language utterance during a conversation between a user and the system;	Alexa Products receive a natural language utterance during a conversation between a user and the system. Examples are receiving the utterances shown in a screenshot of Alexa app's "Review Voice History" page below.

'049 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1. "Each line of the script spoken by the user is known as an utterance. The utterances are then mapped to an intent. When the user speaks one of them the associated intent will be invoked." *Designing for Conversation – Build for dialog* at Step 2.
[11.c] identify a first model that includes short-term knowledge about the conversation, wherein the short-term knowledge is based on one or more prior natural language utterances received during the conversation;	Alexa Products identify a first model that includes short-term knowledge about the conversation, wherein the short-term knowledge is based on one or more prior natural language utterances received during the conversation. An example is identifying a model that includes short term knowledge based on a prior natural language utterance ("alexa what's the weather in seattle" or "alexa what about portland") received during the conversation, as shown in a screenshot of Alexa app's "Review Voice History" page below.

'049 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "Conversations are not isolated occurrences. They build on previous questions and answers, whether from five minutes or five days ago. When appropriate, voice-first experiences should account for prior context to make the conversation more engaging or relevant to the user." *What is Conversational AI?* "Alexa will soon be able to have more natural conversations with users, thanks to something called 'context carryover.' This means that Alexa will be able to understand follow-up questions and respond appropriately..." *Alexa will soon gain a memory*. "For instance, you could ask "Alexa, how is the weather in Seattle?" and then ask, "What about this weekend?" after Alexa responds." *Alexa will soon gain a memory*. "You can even change the subject, saying "Alexa, how's the weather in Portland?" then "How long does it take to get there?" *Alexa will soon gain a memory*.
[11.d] identify, based on the short-term knowledge, context information for the natural language utterance;	Alexa Products identify, based on the short-term knowledge, context information for the natural language utterance. An example is identifying weather or location based on short-term knowledge as shown in a screenshot of Alexa app's "Review Voice

'049 Patent Claim Language	Exemplary Alexa Products
	History" page below.  Below are more examples. "Everyday conversations take many things into account besides just the words we speak, including why, when, and where we converse. When designing for voice, think about how you can provide context to provide a delightful conversation." *What is Conversational AI?* "Today, Alexa analyzes the semantic content of utterances according to the categories domain, intent, and slot." *How Alexa is Learning.* "The system consists of two parts, Gupta explained: (1) an encoder that produces summary vectors from conversation data, and (2) a decoder that calculates a confidence score for each slot. The two together "learn" slot names and all of their possible values to create embeddings — i.e., points in geometric space that represent strings of words — and then leverage a long short-term memory (LSTM) encoder (a neural network that accounts for the sequencing of data) to identify words helpful in determining candidate slot mapping." *AI helps Alexa*.
[11.e] determine, based on the context information, an interpretation of the natural language utterance; and	Alexa Products determine, based on the context information, an interpretation of the natural language utterance. An example is determining an interpretation of the utterance, as shown below in a screenshot of Alexa app's "Review Voice History" page

'049 Patent Claim Language	Exemplary Alexa Products
	below.  Below are more examples. "Today, Alexa analyzes the semantic content of utterances according to the categories domain, intent, and slot." *How Alexa is Learning*. "If, for instance, a customer says, "How far is it to Redmond?" and after the answer follows up by saying, "Find good Indian restaurants there", Alexa should be able to infer that "there" refers to Redmond." *How Alexa is Learning*. "The system consists of two parts, Gupta explained: (1) an encoder that produces summary vectors from conversation data, and (2) a decoder that calculates a confidence score for each slot. The two together "learn" slot names and all of their possible values to create embeddings — i.e., points in geometric space that represent strings of words — and then leverage a long short-term memory (LSTM) encoder (a neural network that accounts for the sequencing of data) to identify words helpful in determining candidate slot mapping." *AI helps Alexa*.

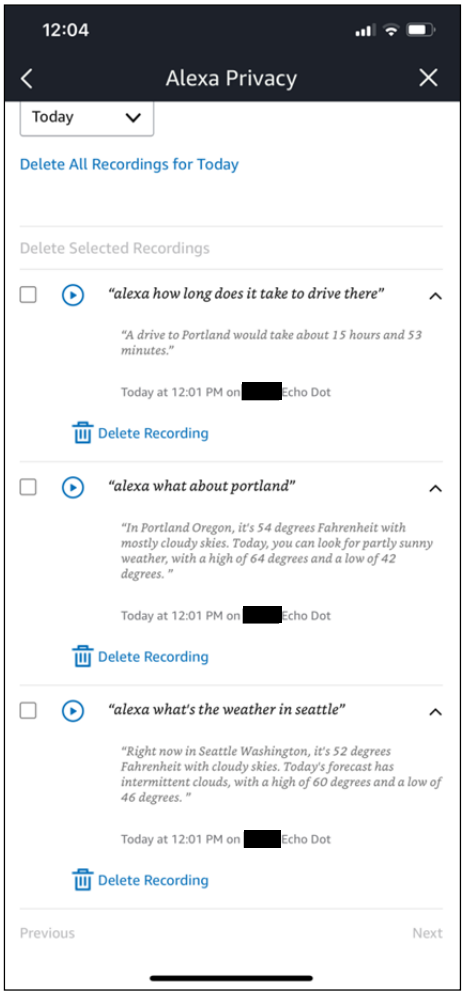
'049 Patent Claim Language	Exemplary Alexa Products
[11.f] generate, based on the interpretation of the natural language utterance, a response to the natural language utterance.	Alexa Products generate, based on the interpretation of the natural language utterance, a response to the natural language utterance. An example is one or more of the responses shown below in a screenshot of Alexa app's "Review Voice History" page below.  Below are more examples. "We are providing this more natural way of engaging with Alexa by adding deep learning models to our spoken language understanding (SLU) pipeline that allows us to carry customers' intent and entities within and across domains (i.e., between weather and traffic)." *Making Alexa More Friction-Free.* "Alexa will soon be able to have more natural conversations with users, thanks to something called 'context carryover.' This means that Alexa will be able to understand follow-up questions and respond appropriately..." *Alexa will soon gain a memory*.

EXHIBIT G

US009626703B2

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

(10) **Patent No.: US 9,626,703 B2**
(45) **Date of Patent: Apr. 18, 2017**

(54) **VOICE COMMERCE**

(71) Applicant: **VOICEBOX TECHNOLOGIES CORPORATION**, Bellevue, WA (US)

(72) Inventor: **Michael R. Kennewick, Sr.**, Bellevue, 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 0 days.

(21) Appl. No.: **14/855,334**

(22) Filed: **Sep. 15, 2015**

(65) **Prior Publication Data**

US 2016/0078504 A1 Mar. 17, 2016

Related U.S. Application Data

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

(51) **Int. Cl.**

G06Q 30/06	(2012.01)
G10L 15/18	(2013.01)
G10L 15/22	(2006.01)

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

CPC ***G06Q 30/0601*** (2013.01); *G10L 15/18* (2013.01); *G10L 2015/223* (2013.01)

(58) **Field of Classification Search**

CPC G06Q 30/06; G06Q 30/0617; G06F 17/2785; G06F 17/3064

(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
EP 1320043 A2 6/2003
(Continued)

OTHER PUBLICATIONS

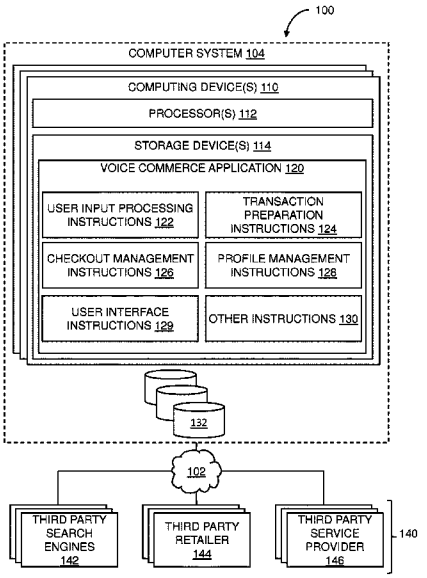
Davis, Z., et al., A Personal Handheld Multi-Modal Shopping Assistant, IEEE, 2006, 9 pages.
(Continued)

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

(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 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.

33 Claims, 6 Drawing Sheets



(58) **Field of Classification Search**
USPC 707/713; 704/10; 379/88.01, 88.1; 705/26.1
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
5,357,596	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
5,617,407	A	4/1997	Bareis
5,633,922	A	5/1997	August
5,634,086	A	5/1997	Rtischev
5,652,570	A	7/1997	Lepkofker
5,675,629	A	10/1997	Raffel
5,696,965	A	12/1997	Dedrick
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
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
5,867,817	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
5,902,347	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
5,983,190	A	11/1999	Trower, II
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
6,076,059	A	6/2000	Glickman
6,078,886	A	6/2000	Dragosh
6,081,774	A	6/2000	deHita
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
6,173,266	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
6,226,612	B1	5/2001	Srenger
6,233,556	B1	5/2001	Teunen
6,233,559	B1	5/2001	Balakrishnan
6,233,561	B1	5/2001	Junqua
6,236,968	B1	5/2001	Kanevsky
6,246,981	B1	6/2001	Papineni
6,246,990	B1	6/2001	Happ
6,266,636	B1	7/2001	Kosaka
6,269,336	B1	7/2001	Ladd
6,272,455	B1	8/2001	Hoshen
6,275,231	B1	8/2001	Obradovich
6,278,377	B1	8/2001	DeLine
6,278,968	B1	8/2001	Franz
6,286,002	B1	9/2001	Axaopoulos
6,288,319	B1	9/2001	Catona
6,292,767	B1	9/2001	Jackson
6,301,560	B1	10/2001	Masters
6,308,151	B1	10/2001	Smith
6,311,159	B1	10/2001	VanTichelen
6,314,402	B1	11/2001	Monaco
6,321,196	B1	11/2001	Franceschi
6,356,869	B1	3/2002	Chapados

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,362,748 B1 3/2002 Huang
6,366,882 B1 4/2002 Bijl
6,366,886 B1 4/2002 Dragosh
6,374,214 B1 4/2002 Friedland
6,377,913 B1 4/2002 Coffman
6,381,535 B1 4/2002 Durocher
6,385,596 B1 5/2002 Wiser
6,385,646 B1 5/2002 Brown
6,393,403 B1 5/2002 Majaniemi
6,393,428 B1 5/2002 Miller
6,397,181 B1 5/2002 Li
6,404,878 B1 6/2002 Jackson
6,405,170 B1 6/2002 Phillips
6,408,272 B1 6/2002 White
6,411,810 B1 6/2002 Maxemchuk
6,411,893 B2 6/2002 Ruhl
6,415,257 B1 7/2002 Junqua
6,418,210 B1 7/2002 Sayko
6,420,975 B1 7/2002 DeLine
6,429,813 B2 8/2002 Feigen
6,430,285 B1 8/2002 Bauer
6,430,531 B1 8/2002 Polish
6,434,523 B1 8/2002 Monaco
6,434,524 B1 8/2002 Weber
6,434,529 B1 8/2002 Walker
6,442,522 B1 8/2002 Carberry
6,446,114 B1 9/2002 Bulfer
6,453,153 B1 9/2002 Bowker
6,453,292 B2 9/2002 Ramaswamy
6,456,711 B1 9/2002 Cheung
6,456,974 B1 9/2002 Baker
6,466,654 B1 10/2002 Cooper
6,466,899 B1 10/2002 Yano
6,470,315 B1 10/2002 Netsch
6,487,494 B2 11/2002 Odinak
6,487,495 B1 11/2002 Gale
6,498,797 B1 12/2002 Anerousis
6,499,013 B1 12/2002 Weber
6,501,833 B2 12/2002 Phillips
6,501,834 B1 12/2002 Milewski
6,505,155 B1 1/2003 Vanbuskirk
6,510,417 B1 1/2003 Woods
6,513,006 B2 1/2003 Howard
6,522,746 B1 2/2003 Marchok
6,523,061 B1 2/2003 Halverson
6,532,444 B1 3/2003 Weber
6,539,348 B1 3/2003 Bond
6,549,629 B2 4/2003 Finn
6,553,372 B1 4/2003 Brassell
6,556,970 B1 4/2003 Sasaki
6,556,973 B1 4/2003 Lewin
6,560,576 B1 5/2003 Cohen
6,560,590 B1 5/2003 Shwe
6,567,778 B1 5/2003 ChaoChang
6,567,797 B1 5/2003 Schuetze
6,567,805 B1 5/2003 Johnson
6,570,555 B1 5/2003 Prevost
6,570,964 B1 5/2003 Murveit
6,571,279 B1 5/2003 Herz
6,574,597 B1 6/2003 Mohri
6,574,624 B1 6/2003 Johnson
6,578,022 B1 6/2003 Foulger
6,581,103 B1 6/2003 Dengler
6,584,439 B1 6/2003 Geilhufe
6,587,858 B1 7/2003 Strazza
6,591,239 B1 7/2003 McCall
6,594,257 B1 7/2003 Doshi
6,594,367 B1 7/2003 Marash
6,598,018 B1 7/2003 Junqua
6,601,026 B2 7/2003 Appelt
6,604,075 B1 8/2003 Brown
6,604,077 B2 8/2003 Dragosh
6,606,598 B1 8/2003 Holthouse
6,611,692 B2 8/2003 Raffel
6,614,773 B1 9/2003 Maxemchuk
6,615,172 B1 9/2003 Bennett
6,622,119 B1 9/2003 Ramaswamy
6,629,066 B1 9/2003 Jackson
6,631,346 B1 10/2003 Karaorman
6,631,351 B1 10/2003 Ramachandran
6,633,846 B1 10/2003 Bennett
6,636,790 B1 10/2003 Lightner
6,643,620 B1 11/2003 Contolini
6,647,363 B2 11/2003 Claassen
6,650,747 B1 11/2003 Bala
6,658,388 B1 12/2003 Kleindienst
6,678,680 B1 1/2004 Woo
6,681,206 B1 1/2004 Gorin
6,691,151 B1 2/2004 Cheyer
6,701,294 B1 3/2004 Ball
6,704,396 B2 3/2004 Parolkar
6,704,576 B1 3/2004 Brachman
6,704,708 B1 3/2004 Pickering
6,707,421 B1 3/2004 Drury
6,708,150 B1 3/2004 Hirayama
6,721,001 B1 4/2004 Berstis
6,721,633 B2 4/2004 Funk
6,721,706 B1 4/2004 Strubbe
6,726,636 B2 4/2004 DerGhazarian
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
6,745,161 B1 6/2004 Arnold
6,751,591 B1 6/2004 Gorin
6,751,612 B1 6/2004 Schuetze
6,754,485 B1 6/2004 Obradovich
6,754,627 B2 6/2004 Woodward
6,757,544 B2 6/2004 Rangarajan
6,757,718 B1 6/2004 Halverson
6,795,808 B1 9/2004 Strubbe
6,801,604 B2 10/2004 Maes
6,801,893 B1 10/2004 Backfried
6,810,375 B1 10/2004 Ejerhed
6,813,341 B1 11/2004 Mahoney
6,816,830 B1 11/2004 Kempe
6,829,603 B1 12/2004 Chai
6,832,230 B1 12/2004 Zilliacus
6,833,848 B1 12/2004 Wolff
6,850,603 B1 2/2005 Eberle
6,856,990 B2 2/2005 Barile
6,865,481 B2 3/2005 Kawazoe
6,868,380 B2 3/2005 Kroeker
6,868,385 B1 3/2005 Gerson
6,871,179 B1 3/2005 Kist
6,873,837 B1 3/2005 Yoshioka
6,877,001 B2 4/2005 Wolf
6,877,134 B1 4/2005 Fuller
6,901,366 B1 5/2005 Kuhn
6,910,003 B1 6/2005 Arnold
6,912,498 B2 6/2005 Stevens
6,915,126 B2 7/2005 Mazzara, Jr.
6,928,614 B1 8/2005 Everhart
6,934,756 B2 8/2005 Maes
6,937,977 B2 8/2005 Gerson
6,937,982 B2 8/2005 Kitaoka
6,941,266 B1 9/2005 Gorin
6,944,594 B2 9/2005 Busayapongchai
6,950,821 B2 9/2005 Faybishenko
6,954,755 B2 10/2005 Reisman
6,959,276 B2 10/2005 Droppo
6,961,700 B2 11/2005 Mitchell
6,963,759 B1 11/2005 Gerson
6,964,023 B2 11/2005 Maes
6,968,311 B2 11/2005 Knockeart
6,973,387 B2 12/2005 Masclet
6,975,993 B1 12/2005 Keiller
6,980,092 B2 12/2005 Turnbull
6,983,055 B2 1/2006 Luo
6,990,513 B2 1/2006 Belfiore
6,996,531 B2 2/2006 Korall
7,003,463 B1 2/2006 Maes
7,016,849 B2 3/2006 Arnold
7,020,609 B2 3/2006 Thrift

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,024,364 B2 4/2006 Guerra
7,027,586 B2 4/2006 Bushey
7,027,974 B1 4/2006 Busch
7,027,975 B1 4/2006 Pazandak
7,035,415 B2 4/2006 Belt
7,036,128 B1 4/2006 Julia
7,043,425 B2 5/2006 Pao
7,054,817 B2 5/2006 Shao
7,058,890 B2 6/2006 George
7,062,488 B1 6/2006 Reisman
7,069,220 B2 6/2006 Coffman
7,072,834 B2 7/2006 Zhou
7,076,362 B2 7/2006 Ohtsuji
7,082,469 B2 7/2006 Gold
7,085,708 B2 8/2006 Manson
7,092,928 B1 8/2006 Elad
7,107,210 B2 9/2006 Deng
7,107,218 B1 9/2006 Preston
7,110,951 B1 9/2006 Lemelson
7,127,395 B1 10/2006 Gorin
7,127,400 B2 10/2006 Koch
7,130,390 B2 10/2006 Abburi
7,136,875 B2 11/2006 Anderson
7,137,126 B1 11/2006 Coffman
7,143,037 B1 11/2006 Chestnut
7,143,039 B1 11/2006 Stifelman
7,146,319 B2 12/2006 Hunt
7,149,696 B2 12/2006 Shimizu
7,165,028 B2 1/2007 Gong
7,170,993 B2 1/2007 Anderson
7,171,291 B2 1/2007 Obradovich
7,174,300 B2 2/2007 Bush
7,177,798 B2 2/2007 Hsu
7,184,957 B2 2/2007 Brookes
7,190,770 B2 3/2007 Ando
7,197,069 B2 3/2007 Agazzi
7,197,460 B1 3/2007 Gupta
7,203,644 B2 4/2007 Anderson
7,206,418 B2 4/2007 Yang
7,207,011 B2 4/2007 Mulvey
7,215,941 B2 5/2007 Beckmann
7,228,276 B2 6/2007 Omote
7,231,343 B1 6/2007 Treadgold
7,236,923 B1 6/2007 Gupta
7,254,482 B2 8/2007 Kawasaki
7,272,212 B2 9/2007 Eberle
7,277,854 B2 10/2007 Bennett
7,283,829 B2 10/2007 Christenson
7,283,951 B2 10/2007 Marchisio
7,289,606 B2 10/2007 Sibal
7,299,186 B2 11/2007 Kuzunuki
7,301,093 B2 11/2007 Sater
7,305,381 B1 12/2007 Poppink
7,321,850 B2 1/2008 Wakita
7,328,155 B2 2/2008 Endo
7,337,116 B2 2/2008 Charlesworth
7,340,040 B1 3/2008 Saylor
7,366,285 B2 4/2008 Parolkar
7,366,669 B2 4/2008 Nishitani
7,376,645 B2 5/2008 Bernard
7,386,443 B1 6/2008 Parthasarathy
7,398,209 B2 7/2008 Kennewick
7,406,421 B2 7/2008 Odinak
7,415,100 B2 8/2008 Cooper
7,415,414 B2 8/2008 Azara
7,421,393 B1 9/2008 DiFabbrizio
7,424,431 B2 9/2008 Greene
7,447,635 B1 11/2008 Konopka
7,451,088 B1 11/2008 Ehlen
7,454,368 B2 * 11/2008 Stillman G06Q 30/06 705/26.35
7,454,608 B2 11/2008 Gopalakrishnan
7,461,059 B2 12/2008 Richardson
7,472,020 B2 12/2008 Brulle-Drews
7,472,060 B1 12/2008 Gorin
7,472,075 B2 12/2008 Odinak
7,477,909 B2 1/2009 Roth
7,478,036 B2 1/2009 Shen
7,487,088 B1 2/2009 Gorin
7,487,110 B2 2/2009 Bennett
7,493,259 B2 2/2009 Jones
7,493,559 B1 2/2009 Wolff
7,502,672 B1 3/2009 Kolls
7,502,738 B2 3/2009 Kennewick
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
7,552,054 B1 6/2009 Stifelman
7,558,730 B2 7/2009 Davis
7,574,362 B2 8/2009 Walker
7,577,244 B2 8/2009 Taschereau
7,606,708 B2 10/2009 Hwang
7,620,549 B2 11/2009 DiCristo
7,634,409 B2 12/2009 Kennewick
7,640,006 B2 12/2009 Portman
7,640,160 B2 12/2009 DiCristo
7,640,272 B2 12/2009 Mahajan
7,672,931 B2 * 3/2010 Hurst-Hiller G06F 17/30654 707/999.003
7,676,365 B2 3/2010 Hwang
7,676,369 B2 3/2010 Fujimoto
7,684,977 B2 3/2010 Morikawa
7,693,720 B2 4/2010 Kennewick
7,697,673 B2 4/2010 Chiu
7,706,616 B2 4/2010 Kristensson
7,729,916 B2 6/2010 Coffman
7,729,918 B2 6/2010 Walker
7,729,920 B2 6/2010 Chaar
7,734,287 B2 6/2010 Ying
7,748,021 B2 6/2010 Obradovich
7,788,084 B2 8/2010 Brun
7,792,257 B1 9/2010 Vanier
7,801,731 B2 9/2010 Odinak
7,809,570 B2 10/2010 Kennewick
7,818,176 B2 10/2010 Freeman
7,831,426 B2 11/2010 Bennett
7,831,433 B1 11/2010 Belvin
7,856,358 B2 12/2010 Ho
7,873,519 B2 1/2011 Bennett
7,873,523 B2 1/2011 Potter
7,873,654 B2 1/2011 Bernard
7,881,936 B2 2/2011 Longe
7,890,324 B2 2/2011 Bangalore
7,894,849 B2 2/2011 Kass
7,902,969 B2 3/2011 Obradovich
7,917,367 B2 3/2011 DiCristo
7,920,682 B2 4/2011 Byrne
7,949,529 B2 5/2011 Weider
7,949,537 B2 5/2011 Walker
7,953,732 B2 5/2011 Frank
7,974,875 B1 7/2011 Quilici
7,983,917 B2 7/2011 Kennewick
7,984,287 B2 7/2011 Gopalakrishnan
8,005,683 B2 8/2011 Tessel
8,015,006 B2 9/2011 Kennewick
8,027,965 B2 9/2011 Takehara
8,032,383 B1 10/2011 Bhardwaj
8,060,367 B2 11/2011 Keaveney
8,069,046 B2 11/2011 Kennewick
8,073,681 B2 12/2011 Baldwin
8,077,975 B2 12/2011 Ma
8,082,153 B2 12/2011 Coffman
8,086,463 B2 12/2011 Ativanichayaphong
8,103,510 B2 1/2012 Sato
8,112,275 B2 2/2012 Kennewick
8,140,327 B2 3/2012 Kennewick
8,140,335 B2 3/2012 Kennewick
8,145,489 B2 3/2012 Freeman
8,150,694 B2 4/2012 Kennewick
8,155,962 B2 4/2012 Kennewick

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,170,867 B2 5/2012 Germain
8,180,037 B1 5/2012 Delker
8,195,468 B2 6/2012 Weider
8,200,485 B1 6/2012 Lee
8,219,399 B2 7/2012 Lutz
8,219,599 B2 7/2012 Tunstall-Pedoe
8,224,652 B2 7/2012 Wang
8,255,224 B2 8/2012 Singleton
8,326,627 B2 12/2012 Kennewick
8,326,634 B2 12/2012 DiCristo
8,326,637 B2 12/2012 Baldwin
8,332,224 B2 12/2012 DiCristo
8,340,975 B1 12/2012 Rosenberger
8,346,563 B1 1/2013 Hjelm
8,370,147 B2 2/2013 Kennewick
8,447,607 B2 5/2013 Weider
8,447,651 B1 5/2013 Scholl
8,452,598 B2 5/2013 Kennewick
8,503,995 B2 8/2013 Ramer
8,509,403 B2 8/2013 Chiu
8,515,765 B2 8/2013 Baldwin
8,527,274 B2 9/2013 Freeman
8,577,671 B1 11/2013 Barve
8,589,161 B2 11/2013 Kennewick
8,620,659 B2 12/2013 DiCristo
8,719,005 B1 5/2014 Lee
8,719,009 B2 5/2014 Baldwin
8,719,026 B2 5/2014 Kennewick
8,731,929 B2 5/2014 Kennewick
8,738,380 B2 5/2014 Baldwin
8,849,652 B2 9/2014 Weider
8,849,670 B2 9/2014 DiCristo
8,849,696 B2 9/2014 Pansari
8,849,791 B1 * 9/2014 Hertschuh G06Q 30/0617 707/713
8,886,536 B2 11/2014 Freeman
8,972,243 B1 3/2015 Strom
8,983,839 B2 3/2015 Kennewick
9,009,046 B1 4/2015 Stewart
9,015,049 B2 4/2015 Baldwin
9,037,455 B1 5/2015 Faaborg
9,070,366 B1 6/2015 Mathias
9,105,266 B2 8/2015 Baldwin
9,171,541 B2 10/2015 Kennewick
9,269,097 B2 2/2016 Freeman
9,305,548 B2 4/2016 Kennewick
9,308,445 B1 4/2016 Merzenich
9,406,078 B2 8/2016 Freeman
9,502,025 B2 11/2016 Kennewick
2001/0039492 A1 11/2001 Nemoto
2001/0041980 A1 11/2001 Howard
2001/0049601 A1 12/2001 Kroeker
2001/0054087 A1 12/2001 Flom
2002/0010584 A1 1/2002 Schultz
2002/0015500 A1 2/2002 Belt
2002/0022927 A1 2/2002 Lemelson
2002/0022956 A1 2/2002 Ukrainczyk
2002/0029186 A1 3/2002 Roth
2002/0029261 A1 3/2002 Shibata
2002/0032752 A1 3/2002 Gold
2002/0035501 A1 3/2002 Handel
2002/0040297 A1 4/2002 Tsiao
2002/0049535 A1 4/2002 Rigo
2002/0049805 A1 4/2002 Yamada
2002/0059068 A1 5/2002 Rose
2002/0065568 A1 5/2002 Silfvast
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
2002/0087312 A1 7/2002 Lee
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
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
2002/0152260 A1 10/2002 Chen
2002/0161646 A1 10/2002 Gailey
2002/0161647 A1 * 10/2002 Gailey G06F 17/30867 705/14.23
2002/0173333 A1 11/2002 Buchholz
2002/0173961 A1 11/2002 Guerra
2002/0184373 A1 12/2002 Maes
2002/0188602 A1 12/2002 Stubler
2002/0198714 A1 12/2002 Zhou
2003/0014261 A1 1/2003 Kageyama
2003/0016835 A1 1/2003 Elko
2003/0046346 A1 3/2003 Mumick
2003/0064709 A1 4/2003 Gailey
2003/0065427 A1 4/2003 Funk
2003/0069734 A1 4/2003 Everhart
2003/0088421 A1 5/2003 Maes
2003/0093419 A1 5/2003 Bangalore
2003/0097249 A1 5/2003 Walker
2003/0110037 A1 6/2003 Walker
2003/0112267 A1 6/2003 Belrose
2003/0115062 A1 6/2003 Walker
2003/0120493 A1 6/2003 Gupta
2003/0135488 A1 7/2003 Amir
2003/0144846 A1 7/2003 Denenberg
2003/0158731 A1 8/2003 Falcon
2003/0161448 A1 8/2003 Parolkar
2003/0174155 A1 9/2003 Weng
2003/0182132 A1 9/2003 Niemoeller
2003/0187643 A1 10/2003 VanThong
2003/0204492 A1 10/2003 Wolf
2003/0206640 A1 11/2003 Malvar
2003/0212550 A1 11/2003 Ubale
2003/0212558 A1 11/2003 Matula
2003/0212562 A1 11/2003 Patel
2003/0225825 A1 12/2003 Healey
2003/0236664 A1 12/2003 Sharma
2004/0006475 A1 1/2004 Ehlen
2004/0010358 A1 1/2004 Oesterling
2004/0025115 A1 2/2004 Sienel
2004/0030741 A1 2/2004 Wolton
2004/0036601 A1 2/2004 Obradovich
2004/0044516 A1 3/2004 Kennewick
2004/0098245 A1 5/2004 Walker
2004/0117179 A1 6/2004 Balasuriya
2004/0117804 A1 6/2004 Scahill
2004/0122674 A1 6/2004 Bangalore
2004/0133793 A1 7/2004 Ginter
2004/0140989 A1 7/2004 Papageorge
2004/0148154 A1 7/2004 Acero
2004/0148170 A1 7/2004 Acero
2004/0158555 A1 8/2004 Seedman
2004/0166832 A1 8/2004 Portman
2004/0167771 A1 8/2004 Duan
2004/0172247 A1 9/2004 Yoon
2004/0172258 A1 9/2004 Dominach
2004/0189697 A1 9/2004 Fukuoka
2004/0193408 A1 9/2004 Hunt
2004/0193420 A1 9/2004 Kennewick
2004/0199375 A1 10/2004 Ehsani
2004/0205671 A1 10/2004 Sukehiro
2004/0243417 A9 12/2004 Pitts
2004/0247092 A1 12/2004 Timmins
2005/0015256 A1 1/2005 Kargman
2005/0021331 A1 1/2005 Huang
2005/0021334 A1 1/2005 Iwahashi
2005/0021470 A1 1/2005 Martin
2005/0021826 A1 1/2005 Kumar
2005/0033574 A1 2/2005 Kim
2005/0033582 A1 2/2005 Gadd
2005/0043940 A1 2/2005 Elder
2005/0080632 A1 4/2005 Endo

(56) **References Cited**

U.S. PATENT DOCUMENTS

2005/0114116 A1 5/2005 Fiedler
2005/0125232 A1 6/2005 Gadd
2005/0131673 A1 6/2005 Koizumi
2005/0137850 A1 6/2005 Odell
2005/0137877 A1 6/2005 Oesterling
2005/0143994 A1 6/2005 Mori
2005/0144013 A1 6/2005 Fujimoto
2005/0144187 A1 6/2005 Che
2005/0149319 A1 7/2005 Honda
2005/0216254 A1 9/2005 Gupta
2005/0234727 A1 10/2005 Chiu
2005/0246174 A1 11/2005 DeGolia
2005/0283364 A1 12/2005 Longe
2005/0283752 A1 12/2005 Fruchter
2006/0041431 A1 2/2006 Maes
2006/0047509 A1 3/2006 Ding
2006/0072738 A1 4/2006 Louis
2006/0074671 A1 4/2006 Farmaner
2006/0100851 A1 5/2006 Schonebeck
2006/0130002 A1 6/2006 Hirayama
2006/0182085 A1 8/2006 Sweeney
2006/0206310 A1 9/2006 Ravikumar
2006/0217133 A1 9/2006 Christenson
2006/0242017 A1 10/2006 Libes
2006/0253281 A1 11/2006 Letzt
2006/0285662 A1 12/2006 Yin
2007/0033005 A1 2/2007 Cristo
2007/0033020 A1 2/2007 Francois
2007/0033526 A1 2/2007 Thompson
2007/0038436 A1 2/2007 Cristo
2007/0038445 A1 2/2007 Helbing
2007/0043569 A1 2/2007 Potter
2007/0043574 A1 2/2007 Coffman
2007/0043868 A1 2/2007 Kumar
2007/0050191 A1 3/2007 Weider
2007/0055525 A1 3/2007 Kennewick
2007/0061067 A1 3/2007 Zeinstra
2007/0061735 A1 3/2007 Hoffberg
2007/0073544 A1 3/2007 Millett
2007/0078708 A1 4/2007 Yu
2007/0078709 A1 4/2007 Rajaram
2007/0078814 A1 4/2007 Flowers
2007/0094003 A1 4/2007 Huang
2007/0112555 A1 5/2007 Lavi
2007/0112630 A1 5/2007 Lau
2007/0118357 A1 5/2007 Kasravi
2007/0124057 A1 5/2007 Prieto
2007/0135101 A1 6/2007 Ramati
2007/0146833 A1 6/2007 Satomi
2007/0162296 A1 7/2007 Altberg
2007/0174258 A1 7/2007 Jones
2007/0179778 A1 8/2007 Gong
2007/0185859 A1 8/2007 Flowers
2007/0186165 A1 8/2007 Maislos
2007/0192309 A1 8/2007 Fischer
2007/0198267 A1 8/2007 Jones
2007/0203736 A1 8/2007 Ashton
2007/0208732 A1 9/2007 Flowers
2007/0214182 A1 9/2007 Rosenberg
2007/0250901 A1 10/2007 McIntire
2007/0265850 A1 11/2007 Kennewick
2007/0266257 A1 11/2007 Camaisa
2007/0276651 A1 11/2007 Bliss
2007/0299824 A1 12/2007 Pan
2008/0034032 A1 2/2008 Healey
2008/0046311 A1 2/2008 Shahine
2008/0059188 A1 3/2008 Konopka
2008/0065386 A1 3/2008 Cross
2008/0065389 A1 3/2008 Cross
2008/0091406 A1 4/2008 Baldwin
2008/0103761 A1 5/2008 Printz
2008/0103781 A1 5/2008 Wasson
2008/0104071 A1 5/2008 Pragada
2008/0109285 A1 5/2008 Reuther
2008/0115163 A1 5/2008 Gilboa
2008/0126091 A1 5/2008 Clark
2008/0133215 A1 6/2008 Sarukkai
2008/0140385 A1 6/2008 Mahajan
2008/0147396 A1 6/2008 Wang
2008/0147410 A1 6/2008 Odinak
2008/0147637 A1 6/2008 Li
2008/0154604 A1 6/2008 Sathish
2008/0162471 A1 7/2008 Bernard
2008/0177530 A1 7/2008 Cross
2008/0189110 A1 8/2008 Freeman
2008/0228496 A1 9/2008 Yu
2008/0235023 A1 9/2008 Kennewick
2008/0235027 A1 9/2008 Cross
2008/0294437 A1 11/2008 Nakano
2008/0294994 A1 11/2008 Kruger
2008/0319751 A1 12/2008 Kennewick
2009/0006077 A1 1/2009 Keaveney
2009/0024476 A1 1/2009 Baar
2009/0052635 A1 2/2009 Jones
2009/0067599 A1 3/2009 Agarwal
2009/0076827 A1 3/2009 Bulitta
2009/0106029 A1 4/2009 DeLine
2009/0117885 A1 5/2009 Roth
2009/0144131 A1 6/2009 Chiu
2009/0144271 A1 6/2009 Richardson
2009/0150156 A1 6/2009 Kennewick
2009/0164216 A1 6/2009 Chengalvarayan
2009/0171664 A1 7/2009 Kennewick
2009/0216540 A1 8/2009 Tessel
2009/0248605 A1 10/2009 Mitchell
2009/0259561 A1* 10/2009 Boys G06Q 10/08 705/26.1
2009/0259646 A1 10/2009 Fujita
2009/0265163 A1* 10/2009 Li G06F 17/2785 704/10
2009/0271194 A1 10/2009 Davis
2009/0273563 A1 11/2009 Pryor
2009/0276700 A1 11/2009 Anderson
2009/0287680 A1* 11/2009 Paek G06F 17/3064
2009/0299745 A1 12/2009 Kennewick
2009/0299857 A1 12/2009 Brubaker
2009/0304161 A1* 12/2009 Pettyjohn G06Q 30/0603 379/88.01
2009/0307031 A1 12/2009 Winkler
2009/0313026 A1 12/2009 Coffman
2009/0319517 A1 12/2009 Guha
2010/0023320 A1 1/2010 Cristo
2010/0029261 A1 2/2010 Mikkelsen
2010/0036967 A1 2/2010 Caine
2010/0049501 A1 2/2010 Kennewick
2010/0049514 A1 2/2010 Kennewick
2010/0057443 A1 3/2010 Cristo
2010/0063880 A1 3/2010 Atsmon
2010/0064025 A1 3/2010 Nelimarkka
2010/0094707 A1 4/2010 Freer
2010/0138300 A1 6/2010 Wallis
2010/0145700 A1 6/2010 Kennewick
2010/0185512 A1 7/2010 Borger
2010/0204986 A1 8/2010 Kennewick
2010/0204994 A1 8/2010 Kennewick
2010/0217604 A1 8/2010 Baldwin
2010/0286985 A1 11/2010 Kennewick
2010/0299142 A1 11/2010 Freeman
2010/0312566 A1 12/2010 Odinak
2010/0331064 A1 12/2010 Michelstein
2011/0022393 A1 1/2011 Waller
2011/0106527 A1* 5/2011 Chiu H04M 3/4936 704/9
2011/0112827 A1* 5/2011 Kennewick G10L 15/18 704/9
2011/0112921 A1 5/2011 Kennewick
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
2011/0307167 A1 12/2011 Taschereau
2012/0022857 A1 1/2012 Baldwin
2012/0046935 A1 2/2012 Nagao

(56) **References Cited**

U.S. PATENT DOCUMENTS

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
2012/0278073	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
2014/0108013	A1	4/2014	Cristo
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
2014/0365222	A1	12/2014	Weider
2015/0019211	A1	1/2015	Simard
2015/0019217	A1	1/2015	Cristo
2015/0019227	A1	1/2015	Anandarajah
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
2015/0228276	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
2016/0078773	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

FOREIGN PATENT DOCUMENTS

EP	1646037	4/2006
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
WO	9946763	9/1999
WO	0021232	1/2000
WO	0046792	1/2000
WO	0171609 A2	9/2001
WO	0178065	10/2001
WO	2004072954	1/2004
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.

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

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.

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. pages 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.

(56) **References Cited**

OTHER PUBLICATIONS

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.
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.

* cited by examiner

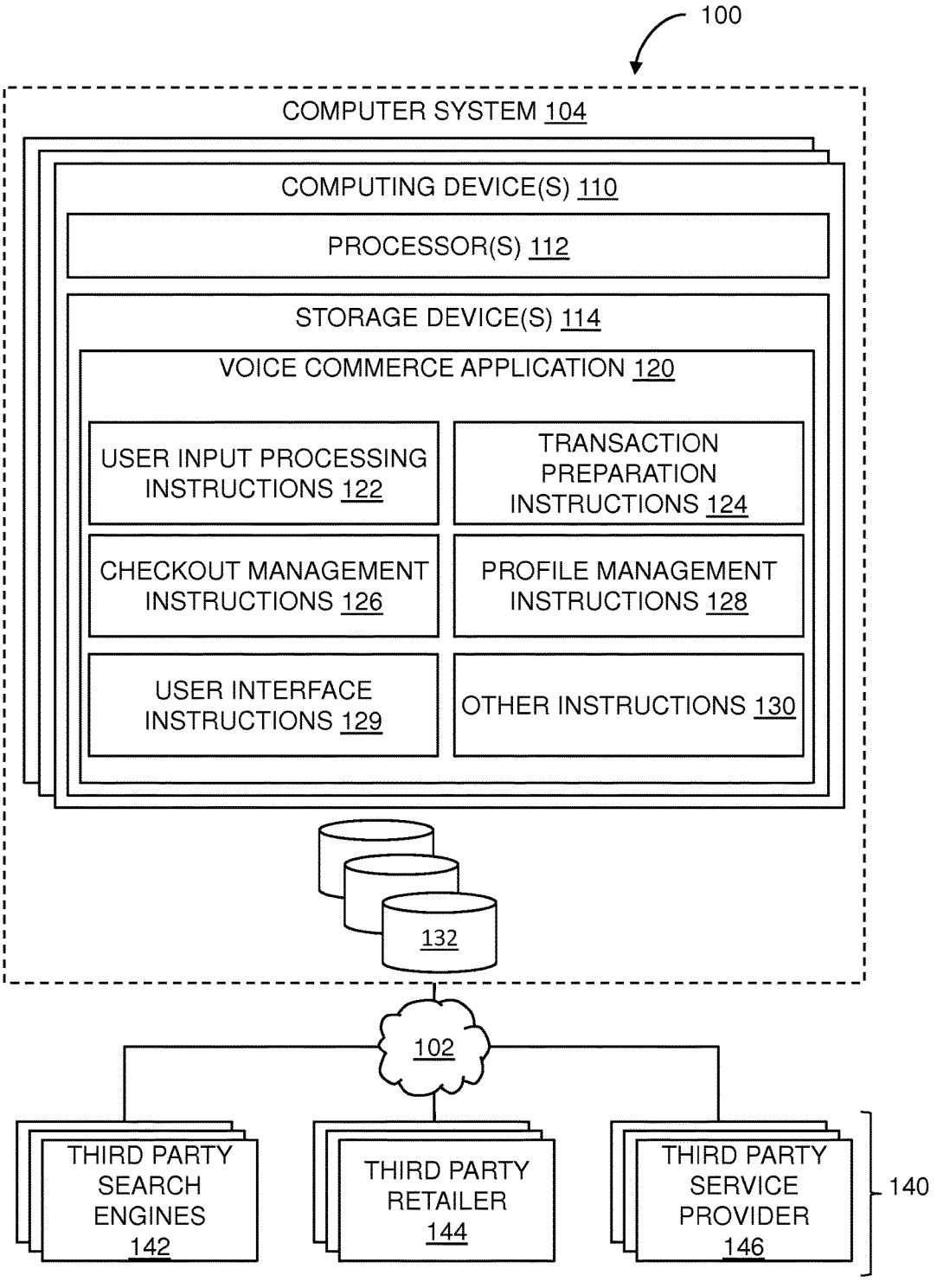

100
COMPUTER SYSTEM 104
COMPUTING DEVICE(S) 110
PROCESSOR(S) 112
STORAGE DEVICE(S) 114
VOICE COMMERCE APPLICATION 120
USER INPUT PROCESSING INSTRUCTIONS 122
TRANSACTION PREPARATION INSTRUCTIONS 124
CHECKOUT MANAGEMENT INSTRUCTIONS 126
PROFILE MANAGEMENT INSTRUCTIONS 128
USER INTERFACE INSTRUCTIONS 129
OTHER INSTRUCTIONS 130
132
102
THIRD PARTY SEARCH ENGINES 142
THIRD PARTY RETAILER 144
THIRD PARTY SERVICE PROVIDER 146
140

FIG. 1

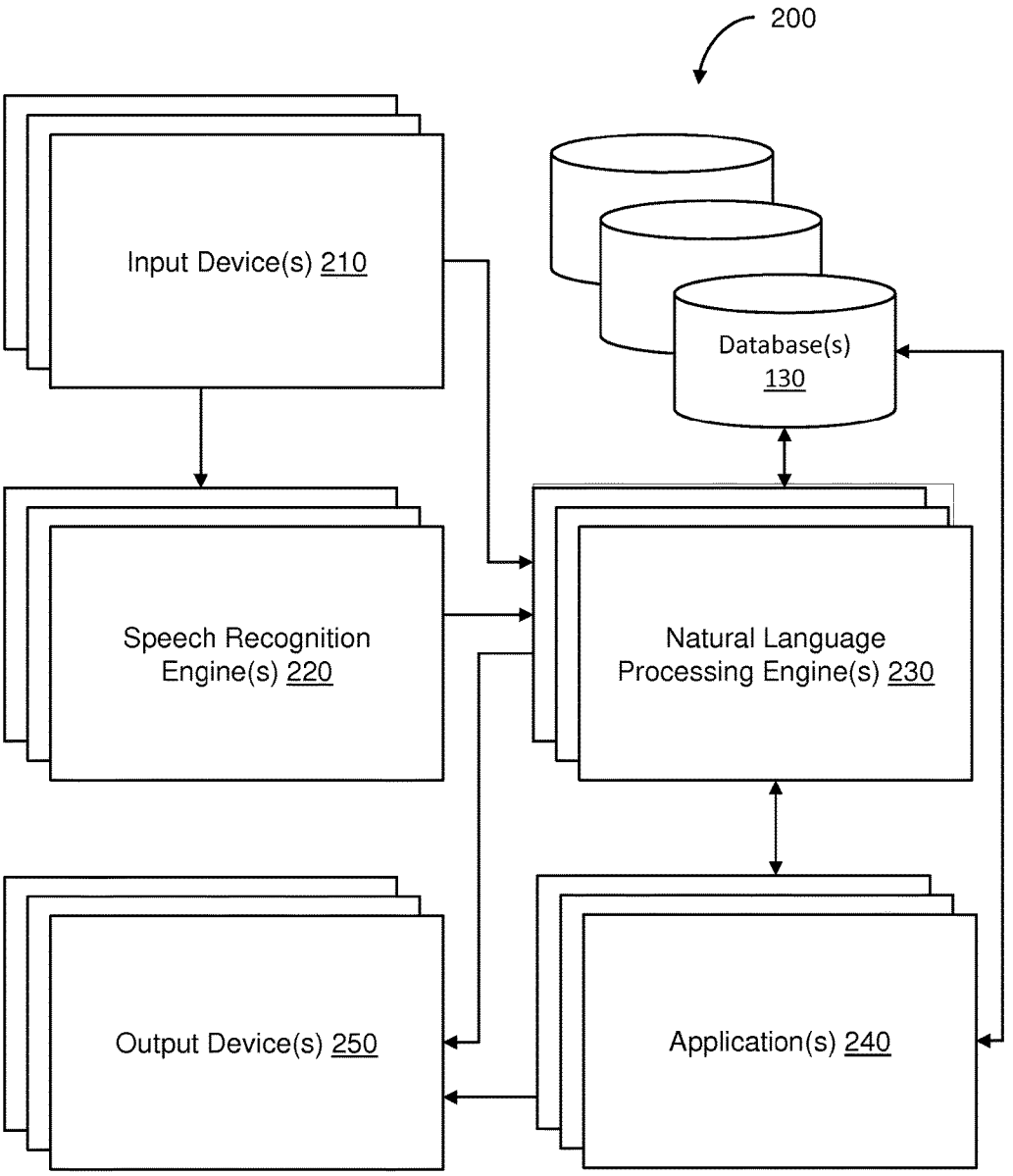

200
Input Device(s) 210
Database(s) 130
Speech Recognition Engine(s) 220
Natural Language Processing Engine(s) 230
Output Device(s) 250
Application(s) 240

FIG. 2

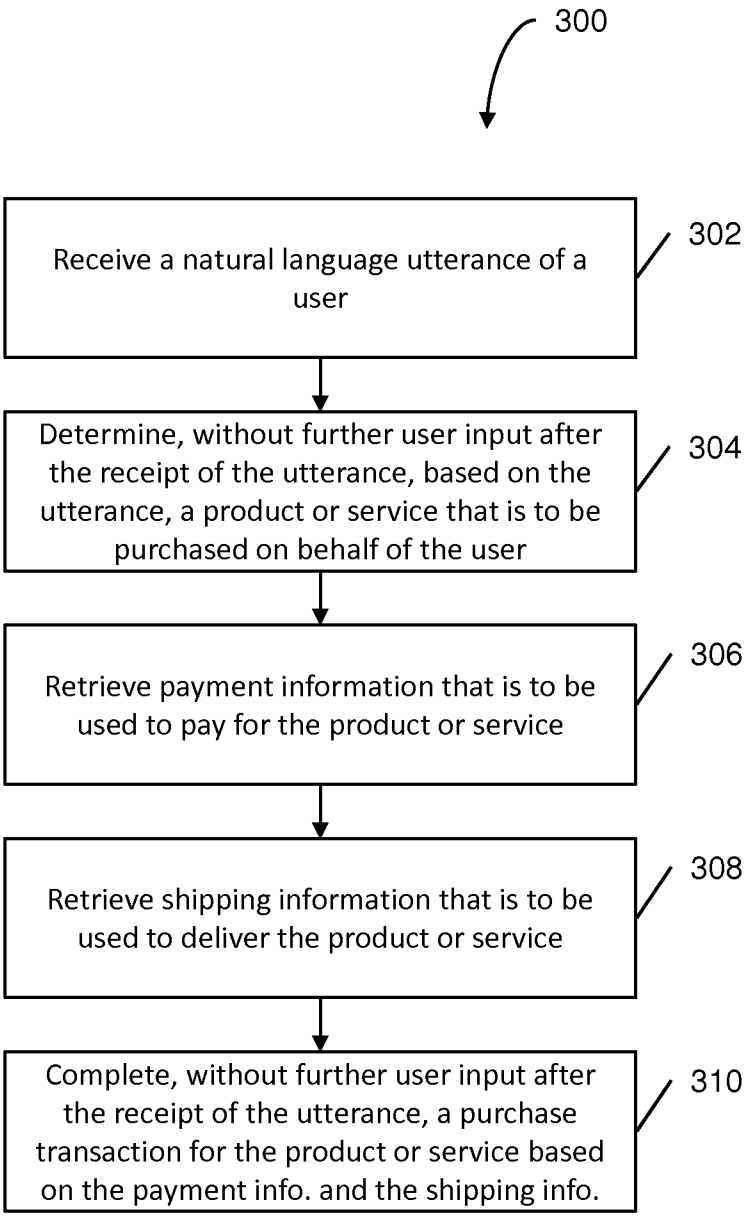

300
Receive a natural language utterance of a user
302
Determine, without further user input after the receipt of the utterance, based on the utterance, a product or service that is to be purchased on behalf of the user
304
Retrieve payment information that is to be used to pay for the product or service
306
Retrieve shipping information that is to be used to deliver the product or service
308
Complete, without further user input after the receipt of the utterance, a purchase transaction for the product or service based on the payment info. and the shipping info.
310

FIG. 3

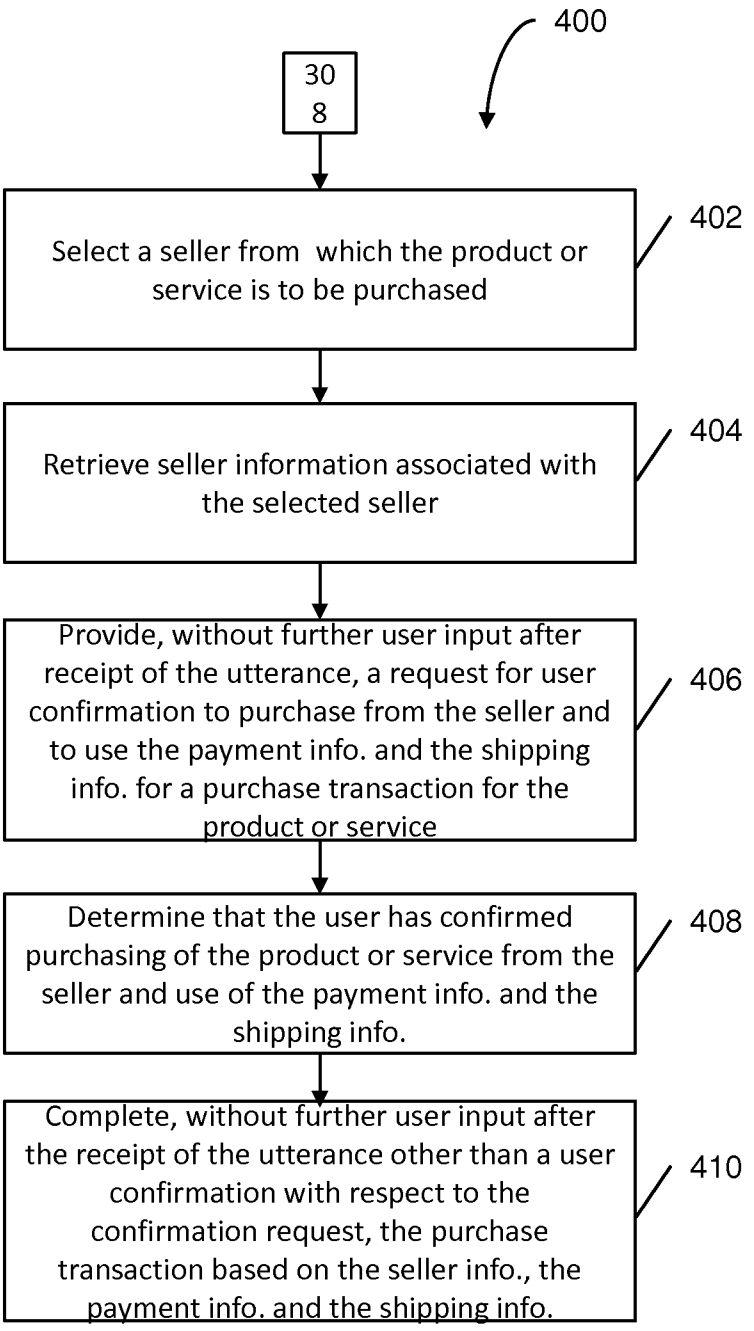

400
308
Select a seller from which the product or service is to be purchased
402
Retrieve seller information associated with the selected seller
404
Provide, without further user input after receipt of the utterance, a request for user confirmation to purchase from the seller and to use the payment info. and the shipping info. for a purchase transaction for the product or service
406
Determine that the user has confirmed purchasing of the product or service from the seller and use of the payment info. and the shipping info.
408
Complete, without further user input after the receipt of the utterance other than a user confirmation with respect to the confirmation request, the purchase transaction based on the seller info., the payment info. and the shipping info.
410

FIG. 4

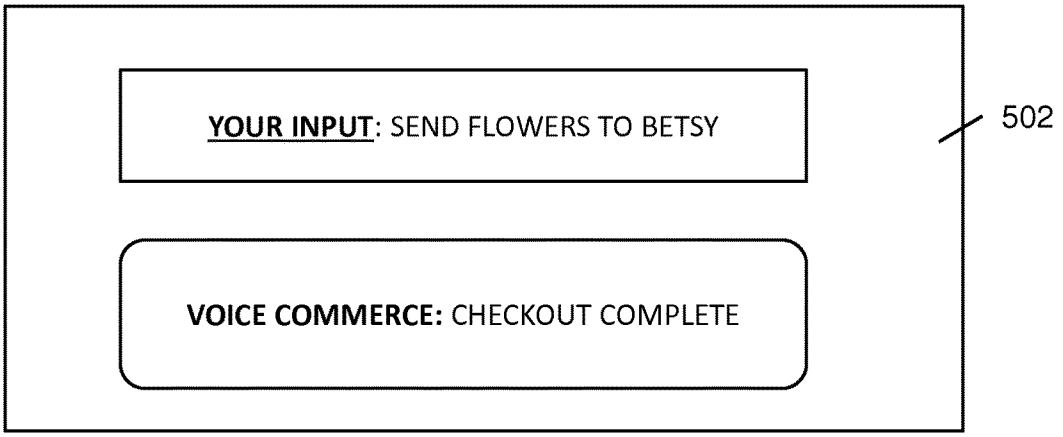


FIG. 5

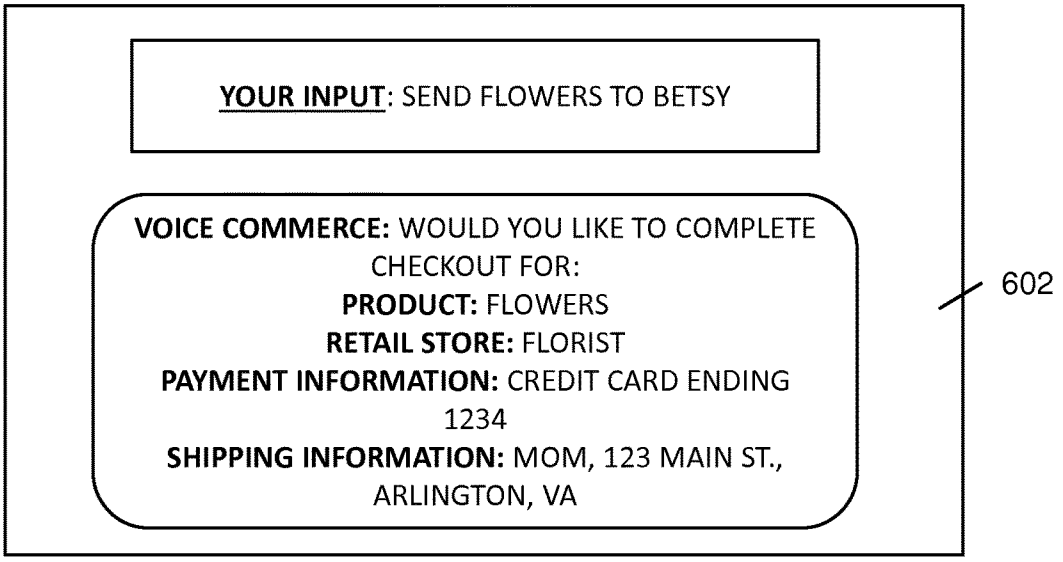


FIG. 6A

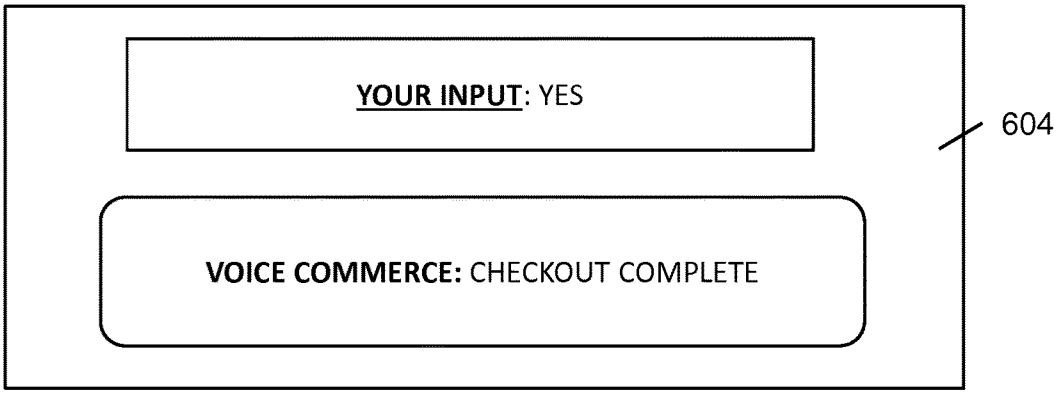


FIG. 6B

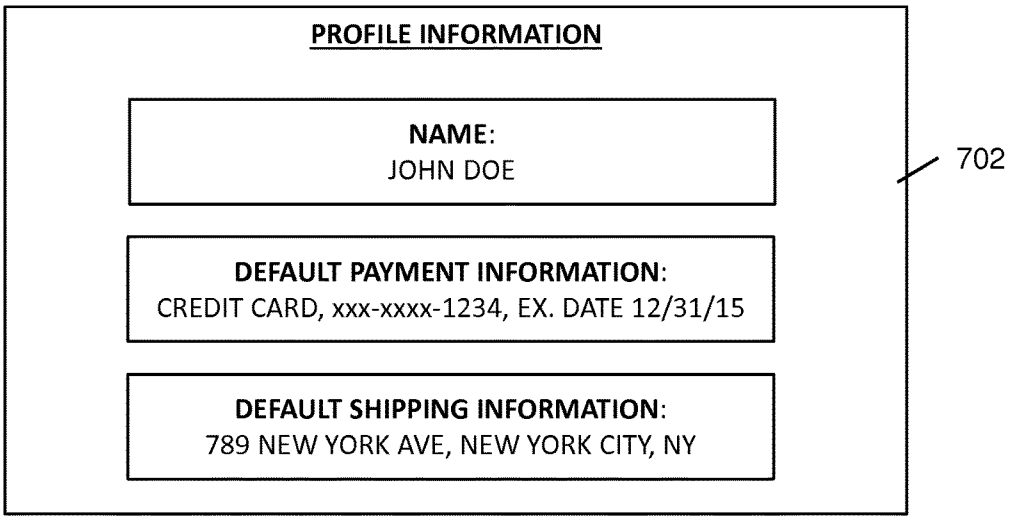

PROFILE INFORMATION
NAME:
JOHN DOE
DEFAULT PAYMENT INFORMATION:
CREDIT CARD, xxx-xxxx-1234, EX. DATE 12/31/15
DEFAULT SHIPPING INFORMATION:
789 NEW YORK AVE, NEW YORK CITY, NY
702

FIG. 7A

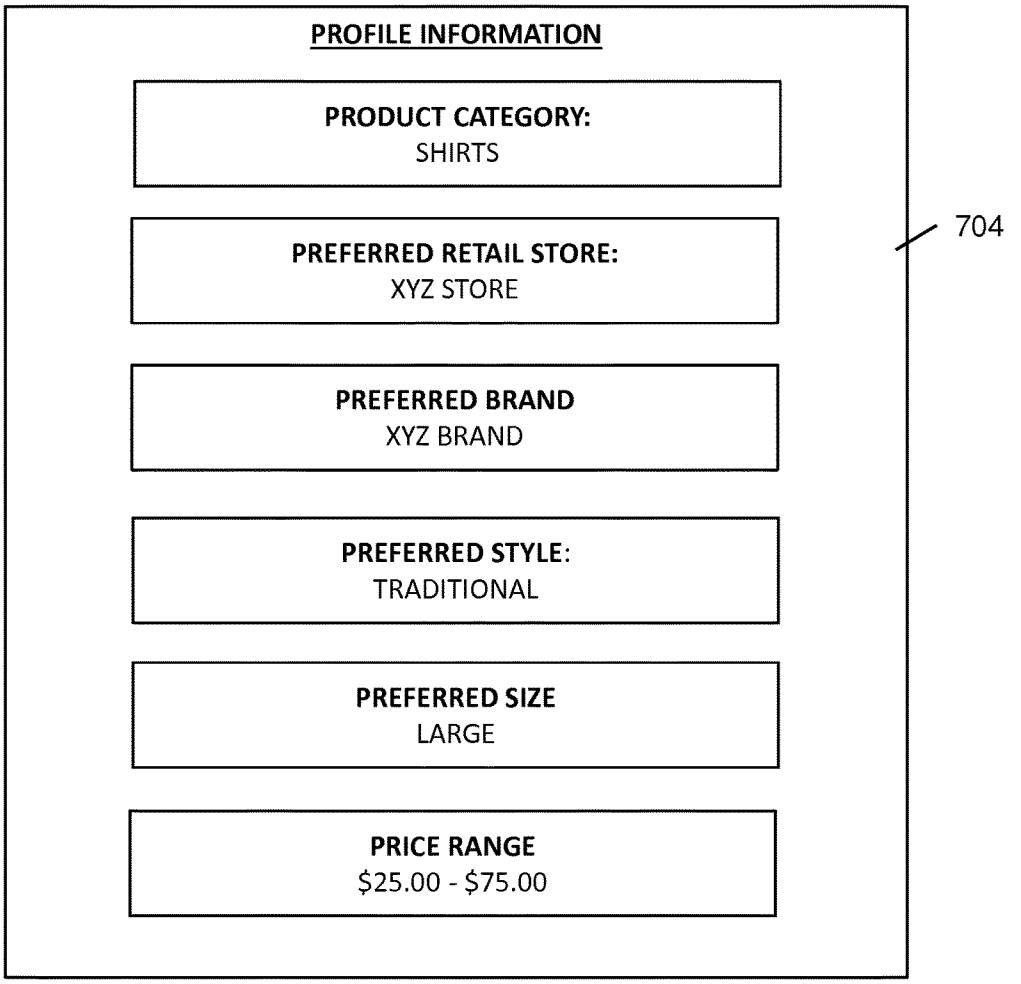

PROFILE INFORMATION
PRODUCT CATEGORY:
SHIRTS
PREFERRED RETAIL STORE:
XYZ STORE
PREFERRED BRAND
XYZ BRAND
PREFERRED STYLE:
TRADITIONAL
PREFERRED SIZE
LARGE
PRICE RANGE
$25.00 - $75.00
704

FIG. 7B

VOICE COMMERCE

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application Ser. No. 62/051,273 filed Sep. 16, 2014 entitled "VOICE COMMERCE", the entirety of which is incorporated herein by reference.

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

FIGS. **7**A and **7**B 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 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 purchase 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 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, 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 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 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 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 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, 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 t 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 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 terms 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 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 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 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**, **6**A, and **5**B.

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 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.

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

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, 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 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. **6**A and **6**B illustrate screenshots **602** and **604** of another 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 **602** of FIG. **6**A, 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 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. **7**A 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. **7**B 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 user input comprising a natural language utterance;

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;

determining, by the computer system, a context based at least on the one or more words or phrases;

identifying, by the computer system, without further user input after the receipt of the user input, a product or service to be purchased on behalf of a user based at least on the determined context;

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

obtaining, by the computer system, without further user input after the receipt of the user input, shipping information with which to deliver the product or service, wherein the shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased; and

completing, by the computer system, without further user input after the receipt of the user input, a purchase transaction for the product or service based on the payment information and shipping information.

2. The method of claim **1**, further comprising:

selecting, by the computer system, a seller from which the product or service is to be purchased; and

obtaining, by the computer system, seller information associated with the seller, wherein completing the purchase transaction comprises completing, without further user input after the receipt of the user input, the purchase transaction based on the seller information, the payment information, and the shipping information.

3. The method of claim **2**, wherein the seller information comprises a price at which the seller will sell the product or service, and wherein completing the purchase transaction comprises completing, without further user input after the receipt of the user input, the purchase transaction based on the price, the payment information, and the shipping information.

4. The method of claim **2**, wherein selecting the seller comprises selecting the seller based on the price at which the seller will sell the product or service.

5. The method of claim **2**, further comprising:

obtaining, by the computer system, user profile information associated with the user, wherein the user profile information indicates a predetermined set of sellers specified by the user,

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

6. The method of claim **2**, 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.

7. The method of claim **1**, further comprising:

obtaining, by the computer system, user profile information associated with the user, wherein the user profile information comprises at least one of default payment information that is to be used to pay for products or services on behalf of the user or default shipping information that is to be used to delivery products or services on behalf of the user,

wherein at least one of: (i) obtaining the payment information comprises selecting the default payment information to pay for the product or service; or (ii) obtaining the shipping information comprises selecting the default shipping information to deliver the product or service, and

wherein completing the purchase transaction comprises completing, without further user input after the receipt of the user input, the purchase transaction based on at least one of the default payment information or the default shipping information.

8. The method of claim **1**, further comprising:

receiving, at the computer system, a previous user input prior to the receipt of the user input, wherein the previous user input is related to the product or service; and

storing, by the computer system, context information associated with the user based on information related to the previous user input,

wherein determining the product or service as at least one product or service to be purchased on behalf of the user comprises determining, without further user input after the receipt of the user input, the product or service based on the natural language utterance and the context information related to the previous user input.

9. The method of claim **8**, further comprising:

performing, by the computer system, speech recognition on the previous user input to recognize one or more words associated with the previous user input; and

identifying, by the computer system, the product or service as at least one product or service related to the previous user input based on the one or more recognized words,

wherein storing the context information comprises storing product information identifying the product or service based on the identification of the product or service as related to the previous user input such that the context information comprises the product information, and

wherein determining the product or service as at least one product or service to be purchased on behalf of the user comprises determining, without further user input after the receipt of the user input, the product or service based on the natural language utterance and the product information.

10. The method of claim **1**, wherein the one or more words or phrases identify at least one of the product or service or a product or service type.

11. The method of claim **1**, wherein the one or more words or phrases by themselves do not indicate the product or service or a product or service type.

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

receiving, at the computer system, a previous user input prior to the receipt of the user input, wherein the previous user input is related to the product or service,

wherein determining the product or service as at least one product or service to be purchased on behalf of the user comprises determining, based on the previous user input, without further user input after the receipt of the user input, the product or service as at least one product or service to be purchased on behalf of the user.

13. The method of claim **1**, further comprising:

obtaining, by the computer system, user profile information associated with the user, wherein the user profile information indicates one or more pre-approved categories of products or services;

determining whether the product or service is associated with at least one of the one or more pre-approved categories;

wherein completing the purchase transaction comprises completing, without further user input after the receipt of the user input, the purchase transaction based on the payment information, the shipping information, and a determination that the product or service is associated with at least one of the one or more pre-approved categories.

14. The method of claim **1**, further comprising:

obtaining, by the computer system, user profile information associated with the user, wherein the user profile information indicates a pre-approved cost threshold;

determining whether a cost related to purchasing the product or service satisfies the pre-approved cost threshold;

wherein completing the purchase transaction comprises completing, without further user input after the receipt of the user input, the purchase transaction based on the payment information, the shipping information, and a determination that the cost related to purchasing product or service satisfies the pre-approved cost threshold.

15. 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 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;

determine a context based at least on the one or more words or phrases;

identify, without further user input after the receipt of the user input, a product or service to be purchased on behalf of a user based at least on the determined context;

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

obtain, without further user input after the receipt of the user input, shipping information with which to deliver the product or service, wherein the shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased; and

complete, without further user input after the receipt of the user input, a purchase transaction for the product or service based on the payment information and shipping information.

16. The system of claim **15**, wherein the one or more physical processors are further caused to:

select a seller from which the product or service is to be purchased; and

obtain seller information associated with the seller,

wherein to complete the purchase transaction, the one or more physical processors are further caused to: complete, without further user input after the receipt of the user input, the purchase transaction based on the seller information, the payment information, and the shipping information.

17. The system of claim **16**, wherein the seller information comprises a price at which the seller will sell the product or service, and wherein to complete the purchase transaction,

the one or more physical processors are further caused to: complete, without further user input after the receipt of the user input, the purchase transaction based on the price, the payment information, and the shipping information.

18. The system of claim **16**, wherein selecting the seller comprises selecting the seller based on the price at which the seller will sell the product or service.

19. The system of claim **16**, 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 specified by the user,

wherein to select the seller, the one or more physical processors are further caused to: select the seller from the predetermined set of user-specified sellers indicated by the user profile information.

20. The system of claim **16**, 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 to select the seller, the one or more physical processors are further caused to: select the seller from the predetermined set of administrator-specified sellers.

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

receive a previous user input prior to the receipt of the user input, wherein the previous user input is related to the product or service; and

store context information associated with the user based on information related to the previous user input,

wherein to identify the product or service as at least one product or service to be purchased on behalf of the user, the one or more physical processors are further caused to: determine, without further user input after the receipt of the user input, the product or service based on the natural language utterance and the context information related to the previous user input.

22. The system of claim **21**, wherein the one or more physical processors are further caused to:

perform speech recognition on the previous user input to recognize one or more words associated with the previous user input; and

identify the product or service as at least one product or service related to the previous user input based on the one or more recognized words,

wherein to store the context information, the one or more physical processors are further caused to: store product information identifying the product or service based on the identification of the product or service as related to the previous user input such that the context information comprises the product information, and wherein to determine the product or service as at least one product or service to be purchased on behalf of the user, the one or more physical processors are further caused to: determine, without further user input after the receipt of the user input, the product or service based on the natural language utterance and the product information.

23. The system of claim **15**, wherein the one or more words or phrases by themselves do not indicate the product or service or a product or service type.

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

receive a previous user input prior to the receipt of the user input, wherein the previous user input is related to the product or service,

wherein to identify the product or service as at least one product or service to be purchased on behalf of the user, the one or more physical processors are further caused to: determine, based on the previous user input, without further user input after the receipt of the user input, the product or service as at least one product or service to be purchased on behalf of the user.

25. 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 user input comprising a natural language utterance;

provide, without further user input after the receipt of the user input, the natural language utterance as an input to a speech recognition engine;

obtain, without further user input after the receipt of the user input, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine;

determine, without further user input after the receipt of the user input, a context based at least on the one or more words or phrases;

identify, without further user input after the receipt of the user input, the product or service to be purchased on behalf of the user based at least on the determined context;

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

obtain, without further user input after the receipt of the user input, shipping information with which to deliver the product or service, wherein the shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased; and

provide, without further user input after the receipt of the user input, a request for user confirmation to use the payment information and the shipping information for a purchase transaction for the product or service.

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

determine whether the user has confirmed use of the payment information and the shipping information for the purchase transaction; and

complete, without further user input after the receipt of the user input other than a user confirmation with respect to the confirmation request, the purchase transaction based on the payment information, the shipping information, and a determination that the user has confirmed use of the payment information and the shipping information.

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

select a seller from which the product or service is to be purchased; and

obtain seller information associated with the seller,

wherein to provide the confirmation request, the one or more physical processors are further caused to: provide, without further user input after the receipt of the user input, 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 the purchase transaction.

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

receive a previous user input prior to the receipt of the user input, wherein the previous user input is related to the product or service; and

store context information associated with the user based on information related to the previous user input,

wherein to identify the product or service as at least one product or service to be purchased on behalf of the user, the one or more physical processors are further caused to: identify, without further user input after the receipt of the user input, the product or service based on the natural language utterance and the context information related to the previous user input.

29. The system of claim 28, wherein the one or more physical processors are further caused to:

perform speech recognition on the previous user input to recognize one or more words associated with the previous user input; and

identify the product or service as at least one product or service related to the previous user input based on the one or more recognized words,

wherein to store the context information, the one or more physical processors are further caused to: store product information identifying the product or service based on the identification of the product or service as related to the previous user input such that the context information comprises the product information, and

wherein to identify the product or service as at least one product or service to be purchased on behalf of the user, the one or more physical processors are further caused to: determine, without further user input after the receipt of the user input, the product or service based on the natural language utterance and the product information.

30. 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 user input comprising a natural language utterance;

provide, without further user input after the receipt of the user input, the natural language utterance as an input to a speech recognition engine;

obtain, without further user input after the receipt of the user input, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine;

identify, without further user input after the receipt of the user input, a context based at least on the one or more words or phrases;

determine, without further user input identifying a product or service after the receipt of the user input, a first product or service to be purchased on behalf of a user based at least on the determined context;

obtain, without further user input identifying payment information after the receipt of the user input, first payment information with which to pay for the product or service;

obtain, without further user input identifying shipping information after the receipt of the user input, first shipping information with which to deliver the product or service, wherein the first shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased; and

complete, without further user input identifying a product or service, payment information, or shipping information after the receipt of the user input, a purchase transaction for the first product or service based on the first payment information and the first shipping information.

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

identifying, by the computer system, a recipient of the product or service based on the one or more words or phrases, and wherein obtaining the shipping information comprises:

obtaining, by the computer system, an address of the recipient.

32. The method of claim 31, wherein identifying the recipient comprises:

associating, by the computer system, at least one word, other than a name of the recipient, from among the one or more words or phrases, that identifies the recipient.

33. The method of claim 31, wherein obtaining the address of the recipient comprises:

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

* * * * *

EXHIBIT H

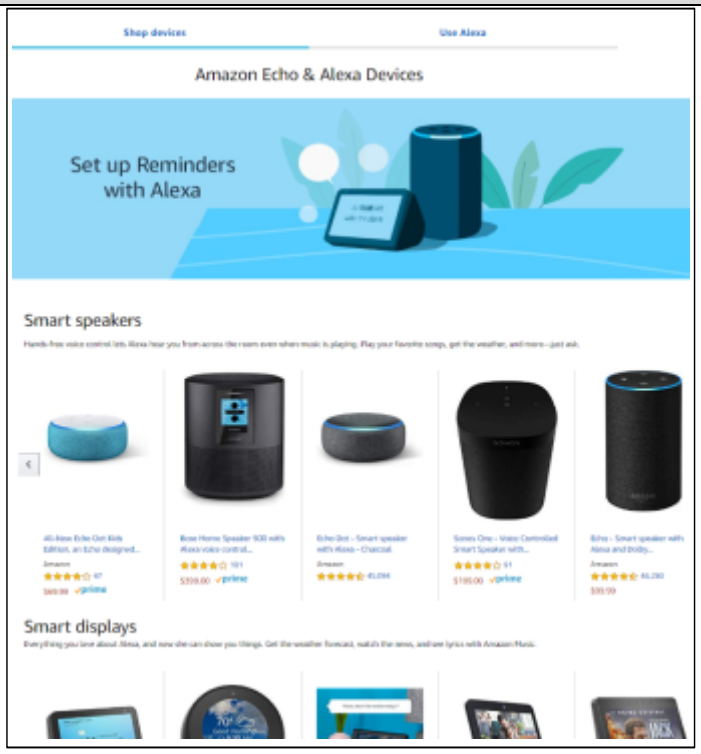
U.S. Pat. No. 9,626,703 ("the '703 patent")
Infringement Chart for Claim 30

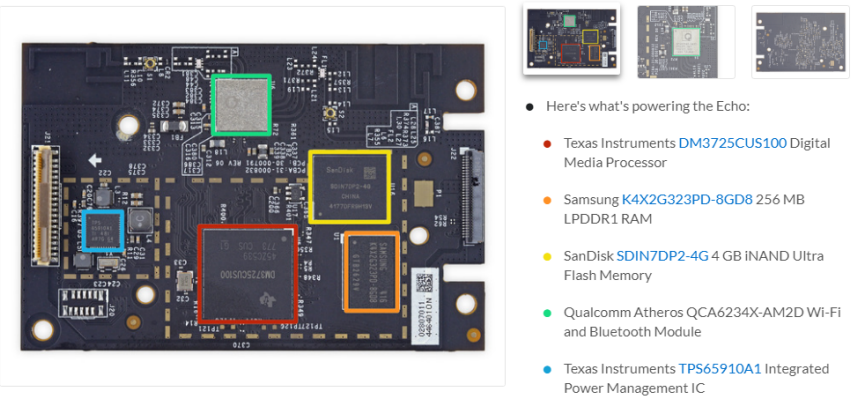
The following chart provides exemplary infringement allegations with respect to an exemplary claim of the '703 patent based on Plaintiff's investigation to date. Plaintiff reserves the right to provide modified and/or more detailed allegations (including with additional claims, products, and evidence) in accordance with its discovery obligations in view of any applicable scheduling order. This chart uses "Alexa Products" to refer to the full scope of instrumentalities identified in the complaint. Any discussion of specific Alexa Products in the following chart is exemplary.

Exemplary information relevant to infringement:

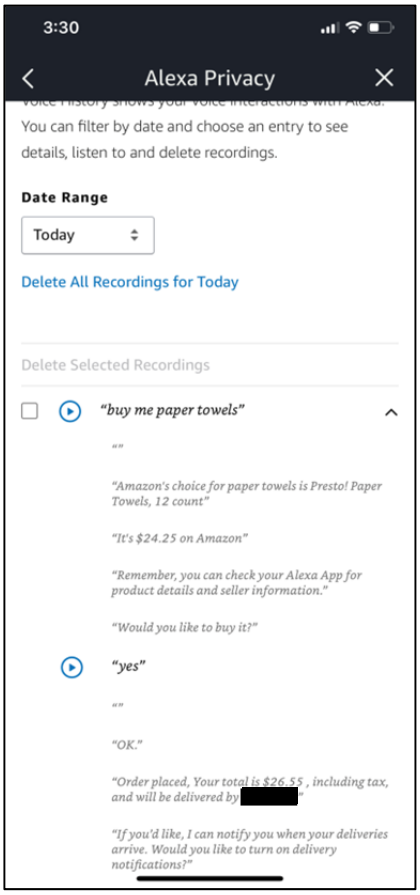
1. *Place Orders with Alexa*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201807210 (last visited Jul. 22, 2019)
2. Taylor Martin, *How to shop with Alexa on Prime Day*, CNET, Jul. 15, 2018, https://www.cnet.com/how-to/how-to-shop-with-alexa-prime-day/ ("*Prime Day*")
3. Richard Baguley & Colin McDonald, *Appliance Science: Alexa, how does Alexa work? The science of the Amazon Echo*, CNET, Aug. 4, 2016, https://www.cnet.com/news/appliance-science-alexa-how-does-alexa-work-the-science-of-amazons-echo/ ("*Appliance Science*")
4. *Audio Hardware Configurations*, Alexa Voice Service, developer.amazon.com, https://developer.amazon.com/docs/alexa-voice-service/audio-hardware-configurations.html (last visited Jul. 22, 2019)
5. *Designing for Conversation – Build for dialog*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/ask-resources/guided/conversational-design-workshop#/lessons/gDcQVF9nOFOtNljrELzARIGN4HuD7LNh?_k=q4qp1a (last visited Jul. 22, 2019)
6. *Amazon Echo & Alexa Devices* - Amazon Official Site, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/?ie=UTF8&node=9818047011&ref_=sv_devicesubnav_1 (last visited Jul. 22, 2019)
7. *Amazon Echo Teardown*, IFIXIT, Dec. 16, 2014, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953
8. *Alexa and Alexa Device FAQs*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (last visited Jul. 22, 2019)
9. *What is Conversational AI?*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/alexa-skills-kit/conversational-ai (last visited Jul. 22, 2019)
10. Chaim Gartenberg, *How to shop on Amazon using Alexa voice controls*, The Verge, Jul. 10, 2017, https://www.theverge.com/2017/7/10/15947672/amazon-alexa-voice-controls-shopping-prime-echo-how-to ("*How to Shop*")

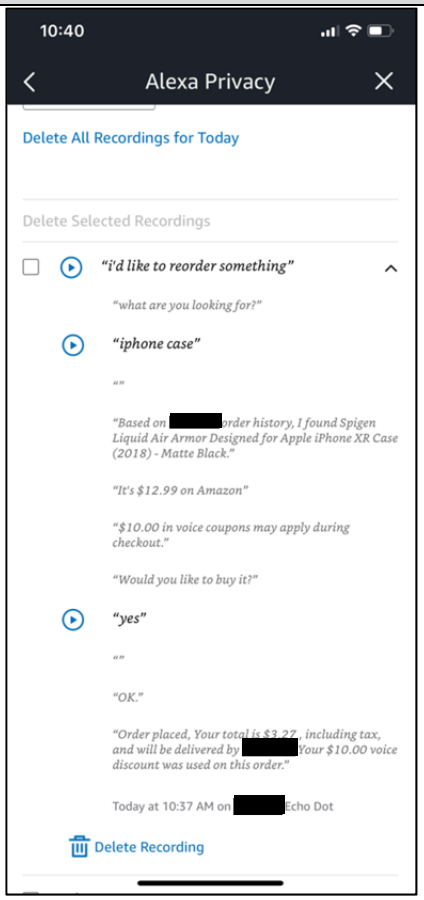
'703 Patent Claim Language	**Exemplary Alexa Products**
30. [pr] A system for providing voice commerce, the system comprising:	Alexa Products include a system for providing voice commerce. For example, Alexa Products (e.g. Echo, Echo Dot, Alexa apps, and Alexa cloud) include Alexa.

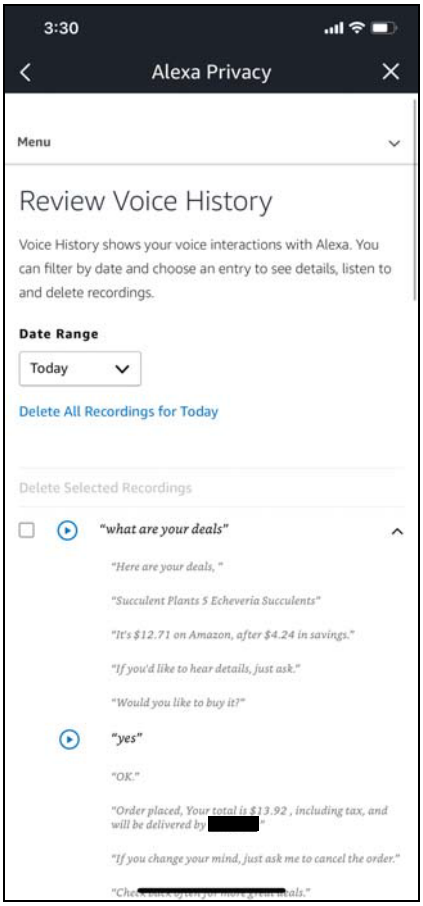
'703 Patent Claim Language	Exemplary Alexa Products
	 E.g. *Amazon Echo & Alexa Devices*. "With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?* Alexa Products include Amazon's Alexa cloud. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1. "You can ask Alexa to place orders for Prime-eligible physical products. On supported devices, you can also ask Alexa to place orders for music." *Place Orders with Alexa*. When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*.
[30.a] one or more physical processors programmed with computer program instructions which, when executed, cause the one or more physical processors to:	Alexa Products include one or more physical processors programmed with computer program instructions which, when executed, cause the one or more physical processors to perform the remaining operations recited in the claim. For example, the Echo has the physical processors shown below:

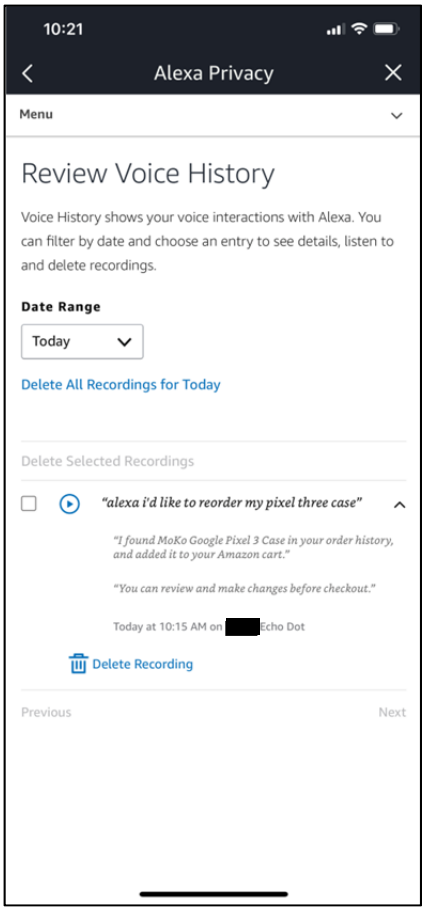
'703 Patent Claim Language	Exemplary Alexa Products
	Step 10  Here's what's powering the Echo: Texas Instruments DM3725CUS100 Digital Media Processor Samsung K4X2G323PD-8GD8 256 MB LPDDR1 RAM SanDisk SDIN7DP2-4G 4 GB iNAND Ultra Flash Memory Qualcomm Atheros QCA6234X-AM2D Wi-Fi and Bluetooth Module Texas Instruments TPS65910A1 Integrated Power Management IC E.g. *Amazon Echo Teardown* at Step 10. As another example, Alexa Products include Amazon's Alexa cloud that has one or more physical processors: "When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask "Alexa, play top hits on Amazon Music" we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2.
[30.b] receive a user input comprising a natural language utterance;	Alexa Products receive a user input comprising a natural language utterance. Examples of receiving an utterance are shown in screenshots of Alexa app's "Review Voice History" page below.

'703 Patent Claim Language	Exemplary Alexa Products



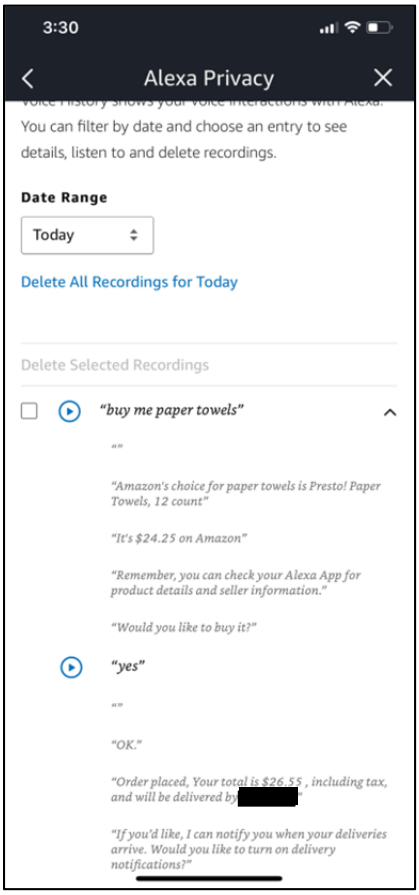


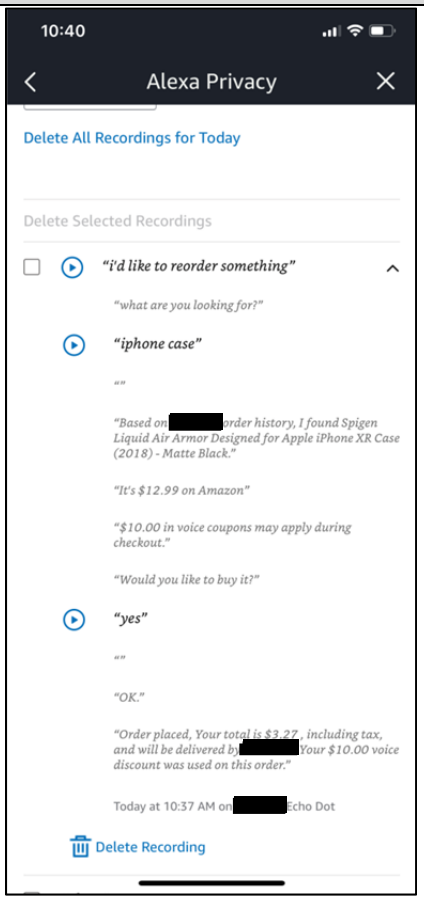


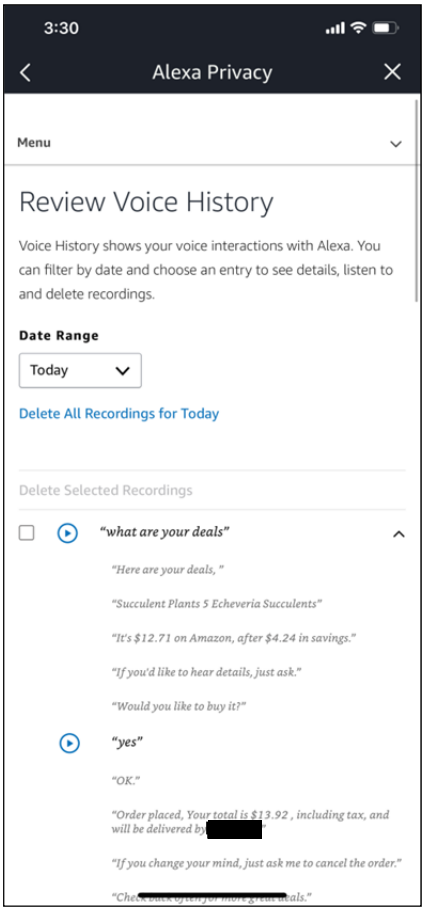


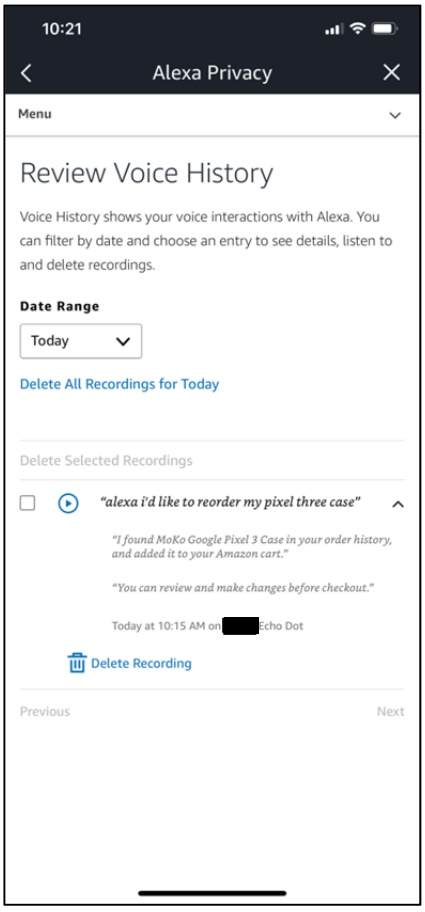
'703 Patent Claim Language	**Exemplary Alexa Products**
	Below are more examples. "Each line of the script spoken by the user is known as an utterance. The utterances are then mapped to an intent. When the user speaks one of them the associated intent will be invoked." *Designing for Conversation – Build for dialog* at Step 2. "You can ask Alexa to place orders for Prime-eligible physical products. On supported devices, you can also ask Alexa to place orders for music." *Place Orders with Alexa.*
[30.c] provide, without further user input after the receipt of the user input, the natural language utterance as an input to a speech recognition engine;	Alexa Products provide, without further user input after the receipt of the user input, the natural language utterance as an input to a speech recognition engine. "When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask 'Alexa, play top hits on Amazon Music' we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2. "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles. "To order something you've not ordered through Amazon before, say something like 'Alexa, order deodorant' . . . Alexa will then perform a search for that item." *Prime Day*.
[30.d] obtain, without further user input after the receipt of the user input, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine;	Alexa Products obtain, without further user input after receipt of the user input, one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine. For example, Alexa Products obtain words or phrases as shown in screenshots of Alexa app's "Review Voice History" page below.

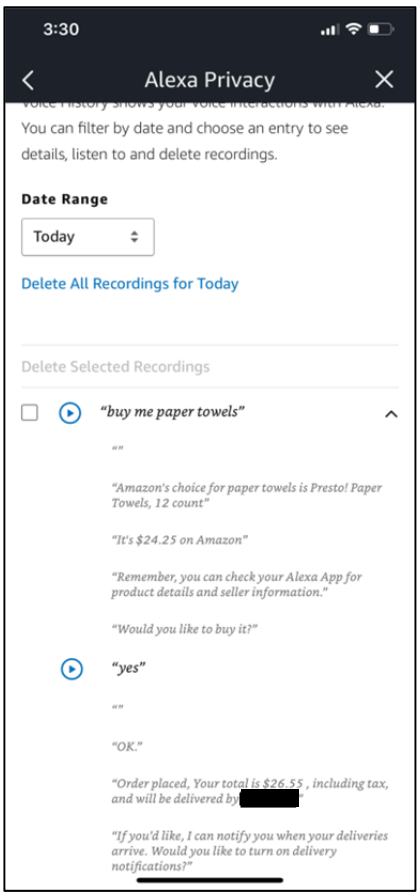
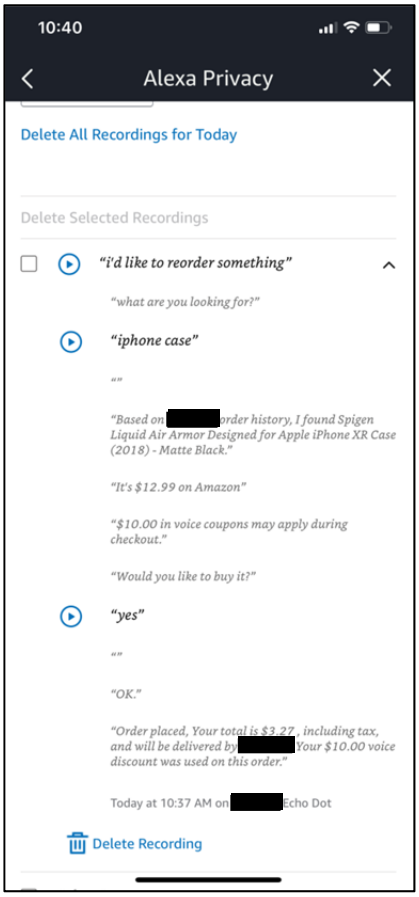
'703 Patent Claim Language	Exemplary Alexa Products

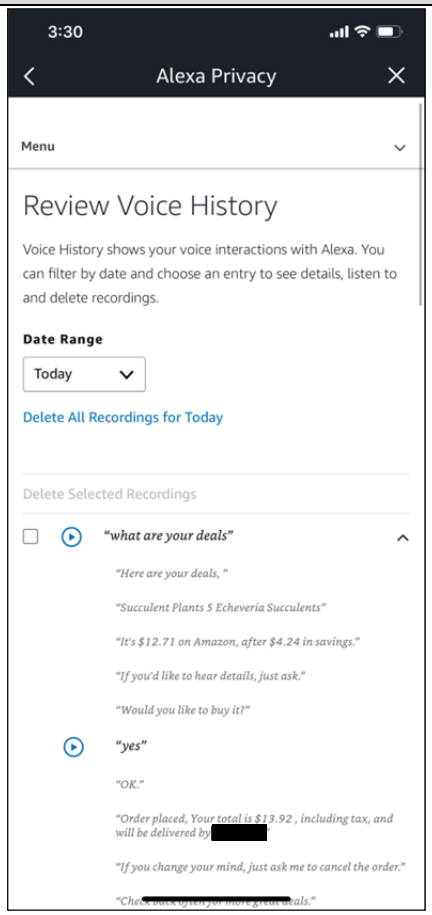
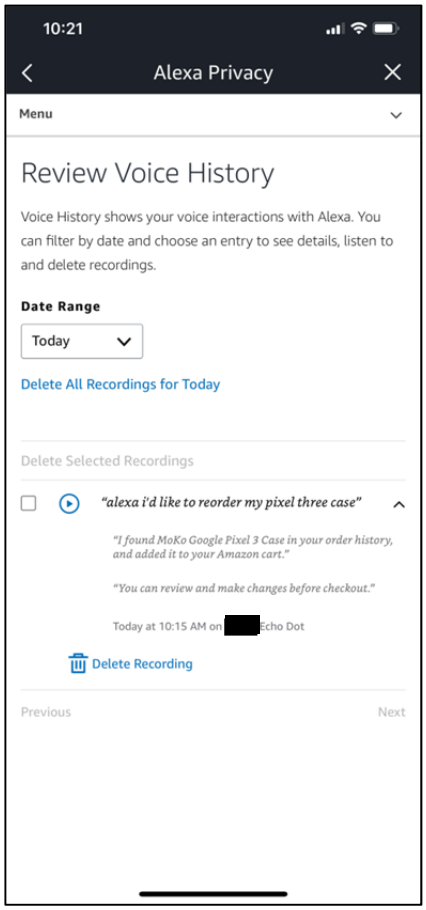




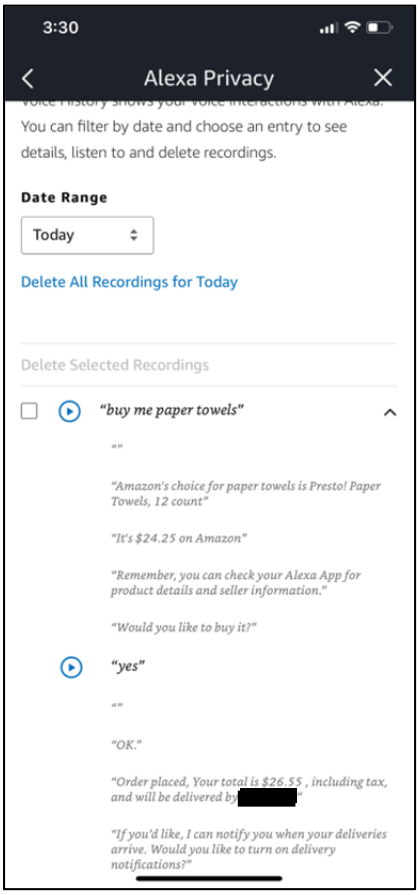


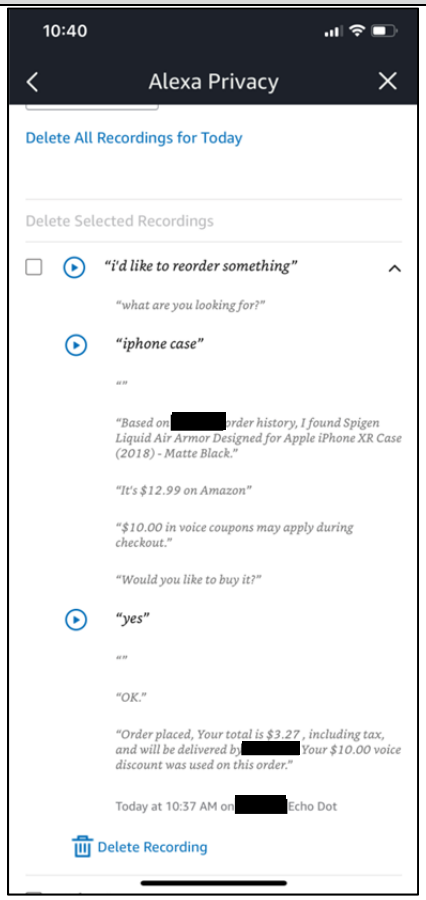


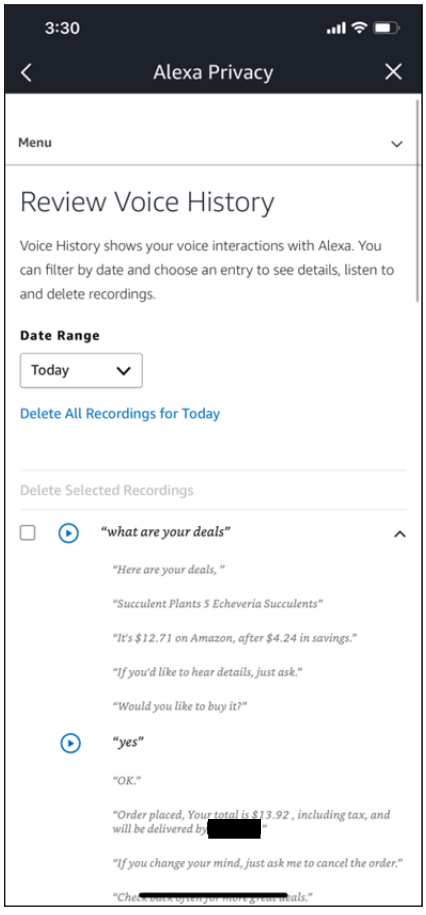
'703 Patent Claim Language	Exemplary Alexa Products
	Below are more examples. "Alexa will then perform a search for that item and offer a few recommendations..." *Prime Day*. "The service that processes this recording is called Alexa Voice Services (AVS). Run by Amazon, this converts the recording into commands that it interprets. It's more than a simple voice-to-text service: it is a fully programmable service that can work with other online services to do a surprising range of things." *Appliance Science*.
[30.e] identify, without further user input after the receipt of the user input, a context based at least on the one or more words or phrases;	Alexa Products identify, without further user input after the receipt of the user input, a context based at least on the one or more words or phrases. Examples are identifying context based on words or phrases shown below in screenshots of Alexa app's "Review Voice History" page.  

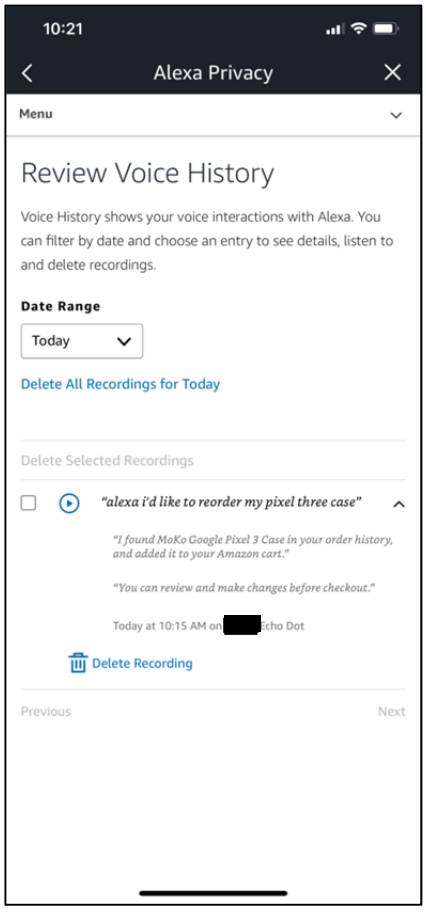
'703 Patent Claim Language	Exemplary Alexa Products
	  Below is another example. "To order something you've not ordered through Amazon before, say something like 'Alexa, order deodorant' . . . Alexa will then perform a search for that item." *Prime Day*.
[30.f] determine, without further user input identifying a product or service after the receipt of the user input, a first product or service to be purchased on behalf of a user based at least on the determined context;	Alexa Products determine, without further user input identifying a product or service after the receipt of the user input, a first product or service to be purchased on behalf of a user based at least on the determined context. Examples are determining a product based on the context as shown in screenshots of Alexa's app's "Review Voice History" page below.

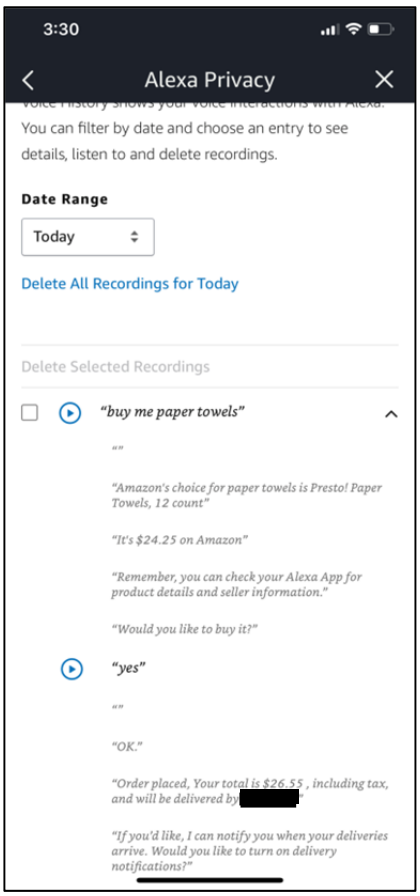
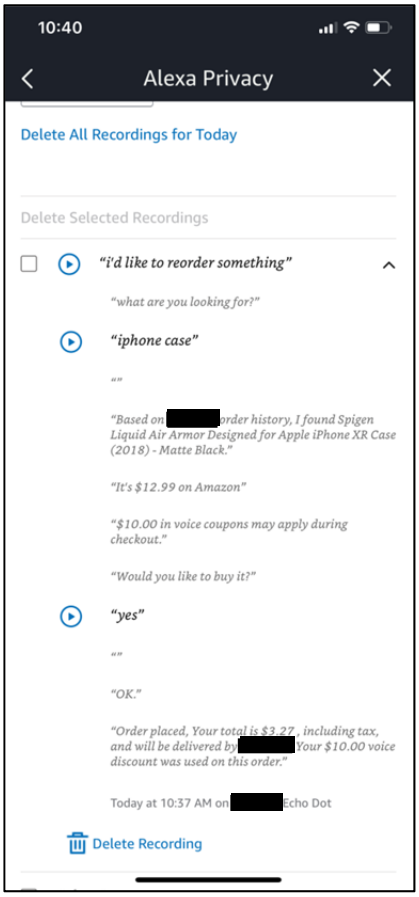
'703 Patent Claim Language	Exemplary Alexa Products



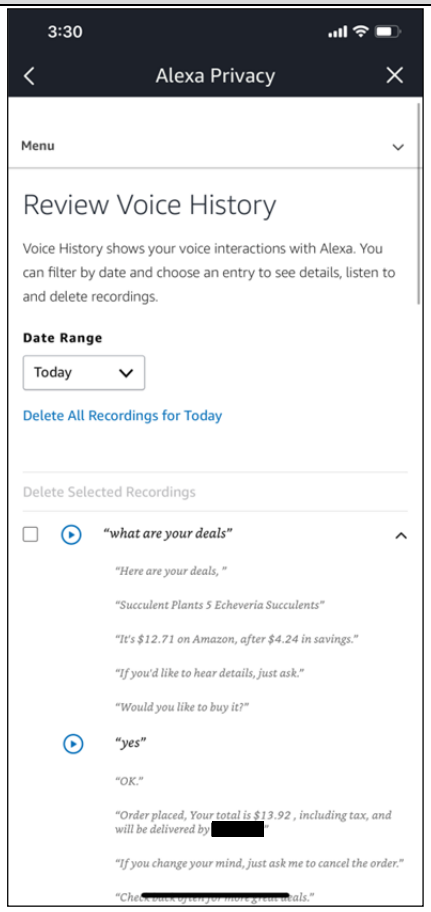




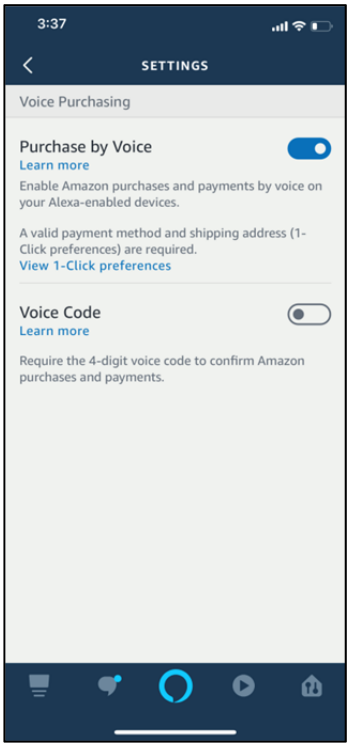


'703 Patent Claim Language	Exemplary Alexa Products
	Below are more examples. "To order something you've not ordered through Amazon before, say something like 'Alexa, order deodorant' . . . Alexa will then perform a search for that item." *Prime Day*. "It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add in your cart." *Prime Day*.
[30.g] obtain, without further user input identifying payment information after the receipt of the user input, first payment information with which to pay for the product or service;	Alexa Products obtain, without further user input identifying payment information after the receipt of the user input, first payment information with which to pay for the product or service. Examples are obtaining payment information, which occurred in the interactions shown in the below screenshots of Alexa app's "Review Voice History" page.  

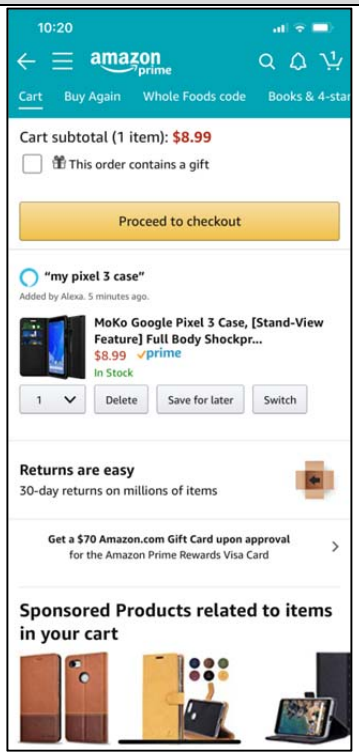
'703 Patent Claim Language	Exemplary Alexa Products



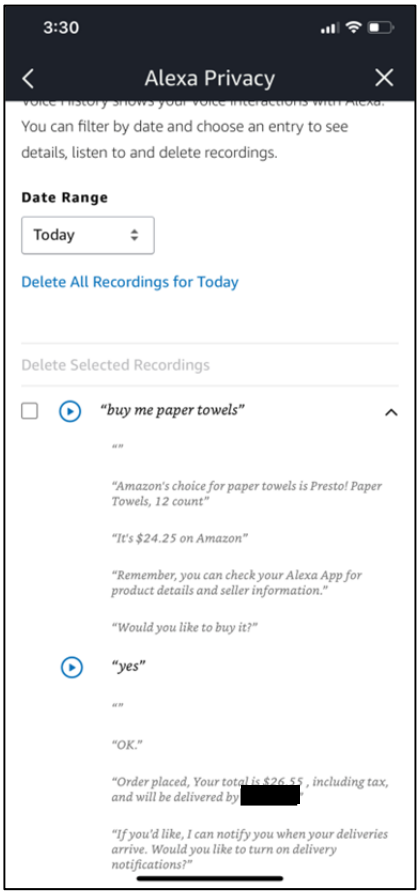
The screenshot below of Alexa's app's "Voice Purchasing" page notes that "a valid payment method...[is] required."

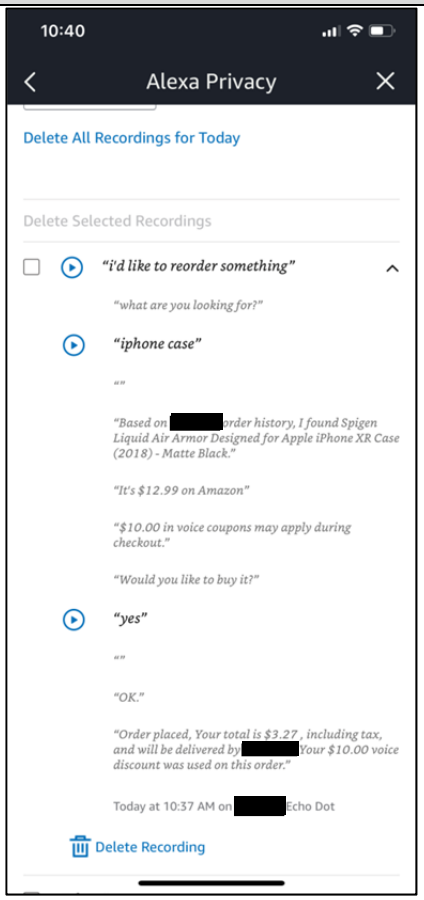


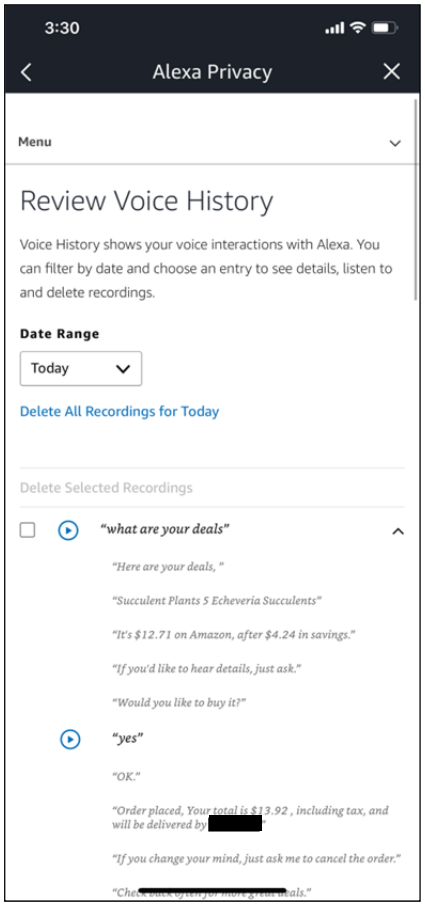
As another example, the Amazon's Shopping app can obtain payment for an item placed into the shopping cart by Alexa.

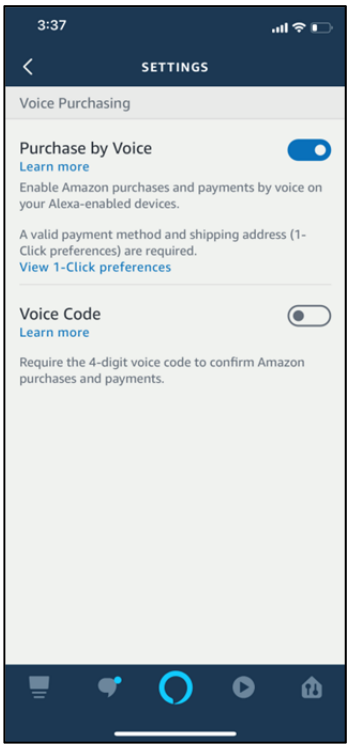
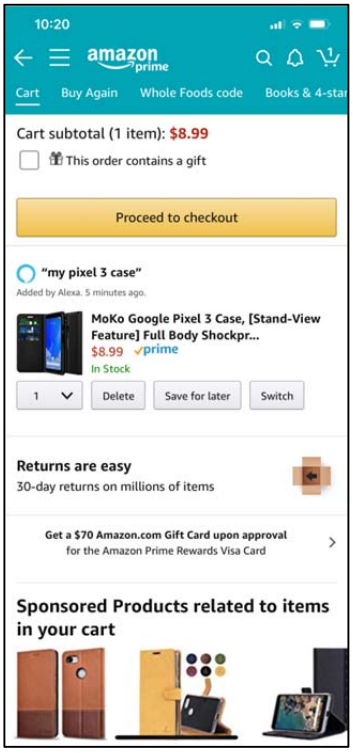
'703 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "To place orders, you need . . . a valid payment method." *Place Orders with Alexa*. "Alexa uses the default . . . payment settings in Your Account." *Place Orders with Alexa.* "You must be an Amazon Prime member with a valid payment method set up prior to ordering something with your voice." *Prime Day*.
[30.h] obtain, without further user input identifying shipping information after the receipt of the user input, first shipping information with which to deliver the product or service, wherein the first shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased; and	Alexa Products obtain, without further user input identifying shipping information after the receipt of the user input, first shipping information with which to deliver the product or service, wherein the first shipping information specifies a name or address of a recipient to which the product or service is to be delivered after the product or service is purchased. Examples are obtaining shipping information, which occurred in the interactions shown in the below screenshots of Alexa app's "Review Voice History" page below.

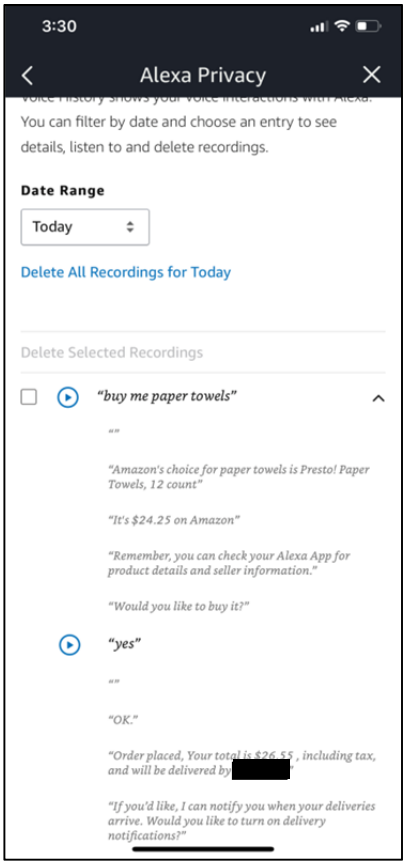
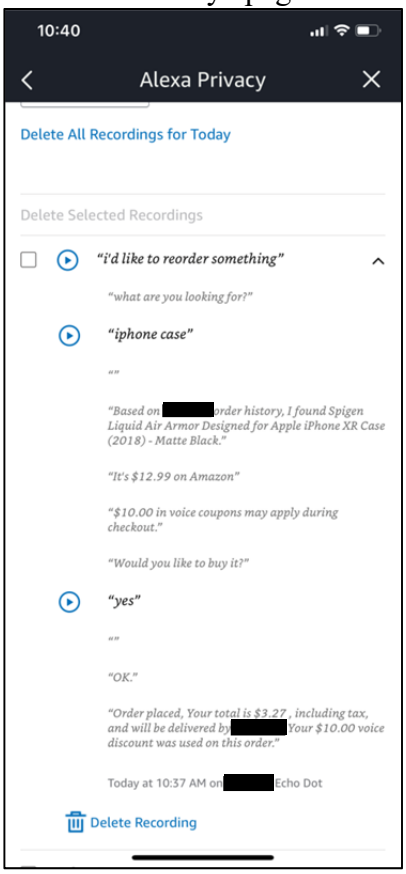
'703 Patent Claim Language	Exemplary Alexa Products

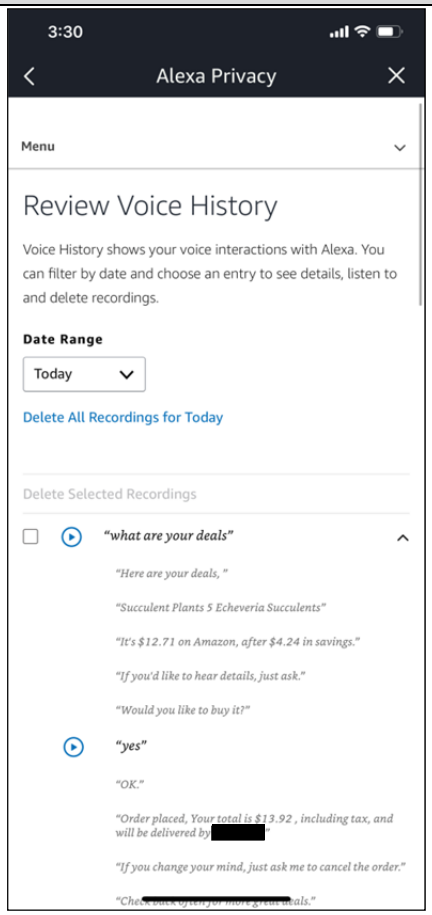






'703 Patent Claim Language	Exemplary Alexa Products
	The screenshot below of Alexa's app's "Voice Purchasing" page notes that "a valid...shipping address...[is] required."  As another example, the Amazon's Shopping app can obtain shipping information for an item placed into the shopping cart by Alexa.  Below is another example. "Alexa uses the default shipping address. . . in Your Account." *Place Orders with Alexa.*

'703 Patent Claim Language	Exemplary Alexa Products
[30.i] complete, without further user input identifying a product or service, payment information, or shipping information after the receipt of the user input, a purchase transaction for the first product or service based on the first payment information and the first shipping information.	Alexa Products complete, without further user input identifying a product or service, payment information, or shipping information after the receipt of the user input, a purchase transaction for the first product or service based on the first payment information and the first shipping information. Examples are completing a purchase transaction, as shown in screenshots of Alexa's app's "Review Voice History" page below. 

'703 Patent Claim Language	Exemplary Alexa Products
	 Another example is completing a purchase transaction from the Amazon's Shopping App for an item placed into the shopping cart by Alexa as shown in the below screenshot of Amazon's Shopping app's "Thank You" page.

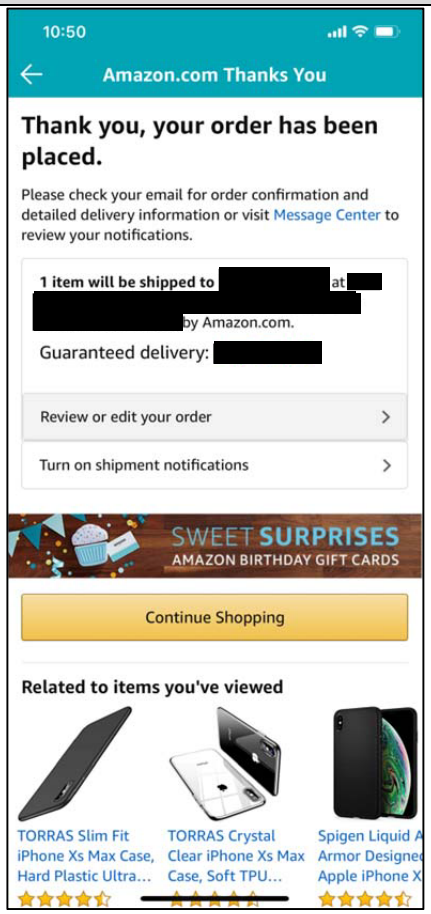
'703 Patent Claim Language	Exemplary Alexa Products
	 Below is another example. "If you want to reorder something you've already purchased from Amazon before, just say, "Alexa, reorder a Swiss Army knife." Alexa will pull up your order history and let you select from the items. You confirm with Alexa which item you want from your history and you're done." *Prime Day*.

EXHIBIT I

US007818176B2

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

(10) **Patent No.: US 7,818,176 B2**
(45) **Date of Patent: Oct. 19, 2010**

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

(75) Inventors: **Tom Freeman**, Mercer Island, WA (US); **Mike Kennewick**, Bellevue, 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 740 days.

(21) Appl. No.: **11/671,526**

(22) Filed: **Feb. 6, 2007**

(65) **Prior Publication Data**

US 2008/0189110 A1 Aug. 7, 2008

(51) **Int. Cl.**

G10L 11/00	(2006.01)
G10L 15/18	(2006.01)
G06F 17/27	(2006.01)
G06Q 30/00	(2006.01)

(52) **U.S. Cl.** **704/270**; 704/9; 704/257; 705/14.49

(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
5,155,743	A	10/1992	Jacobs	375/28
5,274,560	A	12/1993	LaRue	364/444
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
5,488,652	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

(Continued)

FOREIGN PATENT DOCUMENTS

WO	WO 01/78065	10/2001

OTHER PUBLICATIONS

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

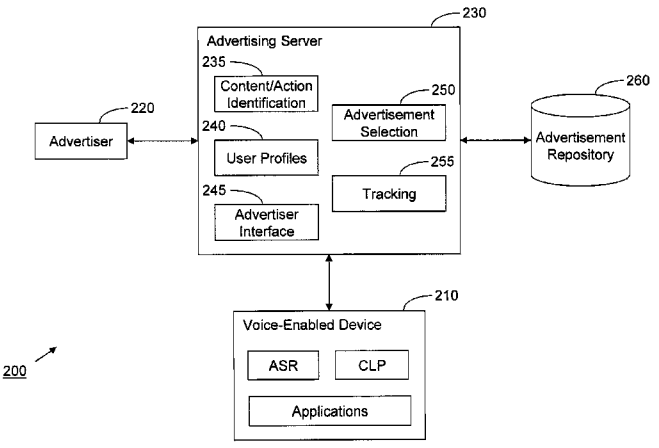
(Continued)

Primary Examiner—Brian L Albertalli
(74) *Attorney, Agent, or Firm*—Pillsbury Winthrop Shaw Pittman LLP

(57) **ABSTRACT**

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

52 Claims, 3 Drawing Sheets



U.S. PATENT DOCUMENTS

5,563,937	A	10/1996	Bruno et al.	379/201
5,590,039	A	12/1996	Ikeda et al.	395/759
5,617,407	A	4/1997	Bareis	369/275.3
5,633,922	A	5/1997	August et al.	379/220
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
5,742,763	A	4/1998	Jones	395/200.3
5,748,841	A	5/1998	Morin et al.	395/2.66
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
5,774,859	A	6/1998	Houser et al.	704/275
5,794,050	A	8/1998	Dahlgren et al.	395/708
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
5,832,221	A	11/1998	Jones	375/200.36
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
5,895,464	A	4/1999	Bhandari et al.	707/3
5,895,466	A	4/1999	Goldberg et al.	707/5
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 et al.	707/1
5,926,784	A	7/1999	Richardson et al.	704/9
5,933,822	A	8/1999	Braden-Harder et al.	707/5
5,953,393	A	9/1999	Culbreth et al.	379/88.25
5,963,894	A	10/1999	Richardson et al.	704/9
5,963,940	A	10/1999	Liddy et al.	707/5
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
6,009,382	A	12/1999	Martino et al.	704/1
6,014,559	A	1/2000	Amin	455/413
6,021,384	A	2/2000	Gorin 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
6,058,187	A	5/2000	Chen	380/21
6,078,886	A	6/2000	Dragosh et al.	704/270
6,081,774	A	6/2000	De Hita et al.	704/9
6,101,241	A	8/2000	Boyce et al.	379/88.01
6,119,087	A	9/2000	Kuhn et al.	704/270
6,134,235	A	10/2000	Goldman et al.	370/352
6,144,667	A	11/2000	Doshi et al.	370/401
6,144,938	A	11/2000	Surace et al.	704/257
6,154,526	A	11/2000	Dahlke et al.	379/88.03
6,160,883	A	12/2000	Jackson et al.	379/230
6,173,279	B1	1/2001	Levin et al.	707/5
6,175,858	B1	1/2001	Bulfer et al.	709/206
6,185,535	B1	2/2001	Hedin et al.	704/270
6,192,110	B1	2/2001	Abella et al.	379/88.01
6,192,338	B1	2/2001	Haszto et al.	704/257
6,195,634	B1	2/2001	Dudemaine et al.	704/231
6,195,651	B1	2/2001	Handel et al.	707/2
6,208,972	B1	3/2001	Grant et al.	704/275
6,219,346	B1	4/2001	Maxemchuk	370/338
6,219,643	B1	4/2001	Cohen et al.	704/257
6,226,612	B1	5/2001	Srenger et al.	704/256
6,233,556	B1	5/2001	Teunen et al.	704/250
6,233,559	B1	5/2001	Balakrishnan	704/275
6,246,981	B1	6/2001	Papineni et al.	704/235
6,272,455	B1	8/2001	Hoshen et al.	704/1
6,288,319	B1	9/2001	Catona	84/609
6,292,767	B1	9/2001	Jackson et al.	704/1
6,308,151	B1	10/2001	Smith	704/235
6,314,402	B1	11/2001	Monaco et al.	704/275
6,366,886	B1	4/2002	Dragosh et al.	704/270.1
6,374,214	B1	4/2002	Friedland et al.	704/235
6,381,535	B1	4/2002	Durocher et al.	701/202
6,385,596	B1	5/2002	Wiser et al.	705/51
6,385,646	B1	5/2002	Brown et al.	709/217
6,393,428	B1	5/2002	Miller et al.	707/102
6,404,878	B1	6/2002	Jackson et al.	379/221.01
6,408,272	B1	6/2002	White et al.	704/270.1
6,411,810	B1	6/2002	Maxemchuk	455/453
6,415,257	B1	7/2002	Junqua et al.	704/275
6,418,210	B1	7/2002	Sayko	379/142.15
6,420,975	B1	7/2002	DeLine et al.	340/815.4
6,430,285	B1	8/2002	Bauer et al.	379/265.01
6,430,531	B1	8/2002	Polish	704/257
6,434,523	B1	8/2002	Monaco	704/257
6,434,524	B1	8/2002	Weber	704/257
6,442,522	B1	8/2002	Carberry et al.	704/257
6,446,114	B1	9/2002	Bulfer et al.	709/206
6,453,153	B1	9/2002	Bowker et al.	455/67.4
6,453,292	B2	9/2002	Ramaswamy et al.	704/235
6,456,711	B1	9/2002	Cheung et al.	379/265.09
6,466,654	B1	10/2002	Cooper et al.	379/88.01
6,466,899	B1	10/2002	Yano et al.	704/1
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
6,513,006	B2	1/2003	Howard et al.	704/257
6,523,061	B1	2/2003	Halverson et al.	709/202
6,532,444	B1	3/2003	Weber	704/257
6,539,348	B1	3/2003	Bond et al.	704/9
6,553,372	B1	4/2003	Brassell et al.	707/5
6,556,970	B1	4/2003	Sasaki et al.	704/257
6,556,973	B1	4/2003	Lewin	704/277
6,560,576	B1	5/2003	Cohen et al.	704/270
6,567,778	B1	5/2003	Chao Chang et al.	704/257
6,567,797	B1	5/2003	Schuetze et al.	707/2
6,570,555	B1	5/2003	Prevost et al.	345/156
6,570,964	B1	5/2003	Murveit et al.	379/67.1
6,574,597	B1	6/2003	Mohri et al.	704/251
6,574,624	B1	6/2003	Johnson et al.	707/5
6,581,103	B1	6/2003	Dengler	709/231
6,594,257	B1	7/2003	Doshi et al.	370/352
6,598,018	B1	7/2003	Junqua	704/251
6,604,077	B2	8/2003	Dragosh et al.	704/270.1
6,611,692	B2	8/2003	Raffel et al.	455/552
6,614,773	B1	9/2003	Maxemchuk	370/337
6,615,172	B1	9/2003	Bennett et al.	704/257
6,629,066	B1	9/2003	Jackson et al.	704/9
6,631,346	B1	10/2003	Karaorman et al.	704/9
6,643,620	B1	11/2003	Contolini et al.	704/270
6,650,747	B1	11/2003	Bala et al.	379/265.06
6,678,680	B1	1/2004	Woo	707/6
6,691,151	B1	2/2004	Cheyer et al.	709/202
6,721,001	B1	4/2004	Berstis	348/231.3
6,721,706	B1	4/2004	Strubbe et al.	704/275
6,735,592	B1	5/2004	Neumann et al.	707/101
6,741,931	B1	5/2004	Kohut et al.	701/209
6,742,021	B1	5/2004	Halverson et al.	709/218
6,757,718	B1	6/2004	Halverson et al.	709/218
6,795,808	B1	9/2004	Strubbe et al.	704/275
6,801,604	B2	10/2004	Maes et al.	379/88.17
6,801,893	B1	10/2004	Backfried et al.	704/257
6,833,848	B1	12/2004	Wolff et al.	345/719
6,856,990	B2	2/2005	Barile et al.	707/10
6,865,481	B2	3/2005	Kawazoe et al.	701/211
6,877,134	B1	4/2005	Fuller et al.	715/500.1
6,901,366	B1	5/2005	Kuhn et al.	704/275
6,937,977	B2	8/2005	Gerson	704/201
6,944,594	B2	9/2005	Busayapongchai et al.	704/275
6,950,821	B2	9/2005	Faybishenko et al.	707/10
6,973,387	B2	12/2005	Masclet et al.	701/211
6,980,092	B2	12/2005	Turnbull et al.	340/425.5
6,990,513	B2	1/2006	Belfiore et al.	709/203

6,996,531	B2	2/2006	Korall et al.	704/270
7,016,849	B2	3/2006	Arnold et al.	704/275
7,020,609	B2	3/2006	Thrift et al.	704/270.1
7,027,975	B1	4/2006	Pazandak et al.	704/9
7,062,488	B1	6/2006	Reisman	707/8
7,082,469	B2	7/2006	Gold et al.	709/231
7,092,928	B1	8/2006	Elad et al.	706/60
7,127,400	B2	10/2006	Koch	704/270.1
7,136,875	B2	11/2006	Anderson et al.	707/104.1
7,137,126	B1	11/2006	Coffman et al.	719/328
7,146,319	B2	12/2006	Hunt	704/254
7,203,644	B2	4/2007	Anderson et al.	704/246
7,228,276	B2	6/2007	Omote et al.	704/243
7,289,606	B2	10/2007	Sibal et al.	379/52
7,301,093	B2	11/2007	Sater et al.	84/615
7,340,040	B1	3/2008	Saylor et al.	379/67.1
7,376,645	B2	5/2008	Bernard	707/3
7,398,209	B2	7/2008	Kennewick et al.	704/255
7,424,431	B2	9/2008	Greene et al.	704/270
7,461,059	B2 *	12/2008	Richardson et al.	707/5
7,487,088	B1	2/2009	Gorin et al.	704/240
7,493,559	B1	2/2009	Wolff et al.	715/727
7,502,738	B2	3/2009	Kennewick et al.	704/257
7,620,549	B2	11/2009	Di Cristo et al.	704/257
7,634,409	B2	12/2009	Kennewick et al.	704/257
7,640,160	B2	12/2009	Di Cristo et al.	704/257
7,693,720	B2	4/2010	Kennewick et al.	704/275
2001/0041980	A1	11/2001	Howard et al.	704/270
2002/0049805	A1	4/2002	Yamada et al.	709/202
2002/0065568	A1	5/2002	Silfvast et al.	700/94
2002/0069059	A1	6/2002	Smith	704/257
2002/0082911	A1	6/2002	Dunn et al.	705/14
2002/0120609	A1	8/2002	Lang et al.	707/1
2002/0124050	A1	9/2002	Middeljans	709/203
2002/0138248	A1	9/2002	Corston-Oliver et al.	704/1
2002/0143535	A1	10/2002	Kist et al.	704/251
2002/0188602	A1	12/2002	Stubler et al.	707/3
2002/0198714	A1	12/2002	Zhou	704/252
2003/0014261	A1	1/2003	Kageyama	704/275
2003/0088421	A1	5/2003	Maes et al.	704/270.1
2003/0097249	A1	5/2003	Walker et al.	704/1
2003/0112267	A1	6/2003	Belrose	345/728
2003/0135488	A1	7/2003	Amir et al.	707/3
2003/0182132	A1	9/2003	Niemoeller	704/275
2004/0025115	A1	2/2004	Sienel et al.	715/513
2004/0044516	A1	3/2004	Kennewick et al.	704/5
2004/0166832	A1	8/2004	Portman et al.	455/412.1
2004/0193420	A1	9/2004	Kennewick et al.	704/257
2004/0199375	A1	10/2004	Ehsani et al.	704/4
2005/0015256	A1	1/2005	Kargman	704/272
2005/0021334	A1	1/2005	Iwahashi	704/240
2005/0021826	A1	1/2005	Kumar	709/232
2005/0033574	A1	2/2005	Kim et al.	704/251
2005/0114116	A1	5/2005	Fiedler	704/201
2005/0246174	A1 *	11/2005	DeGolia	704/270
2006/0206310	A1	9/2006	Ravikumar et al.	704/9
2007/0033005	A1	2/2007	Cristo et al.	704/9
2007/0038436	A1	2/2007	Cristo et al.	704/9
2007/0043574	A1	2/2007	Coffman et al.	704/275
2007/0050191	A1	3/2007	Weider et al.	704/275
2007/0186165	A1 *	8/2007	Maislos et al.	715/728
2007/0214182	A1	9/2007	Rosenberg	707/104.1
2007/0250901	A1	10/2007	McIntire et al.	725/146
2007/0299824	A1	12/2007	Pan et al.	707/3
2008/0091406	A1	4/2008	Baldwin et al.	704/4
2008/0103761	A1 *	5/2008	Printz et al.	704/9
2008/0115163	A1 *	5/2008	Gilboa et al.	725/34
2008/0133215	A1 *	6/2008	Sarukkai	704/2
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
2009/0117885	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
2009/0299745	A1	12/2009	Kennewick et al.	704/257
2010/0023320	A1	1/2010	Di Cristo et al.	704/9
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
2010/0145700	A1	6/2010	Kennewick et al.	704/257

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.

Cheyer 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, [no date], 4 pages.

Weng, Fuliang, et al., "Efficient Lattice Representation and Generation", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, [no date], 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.

* cited by examiner

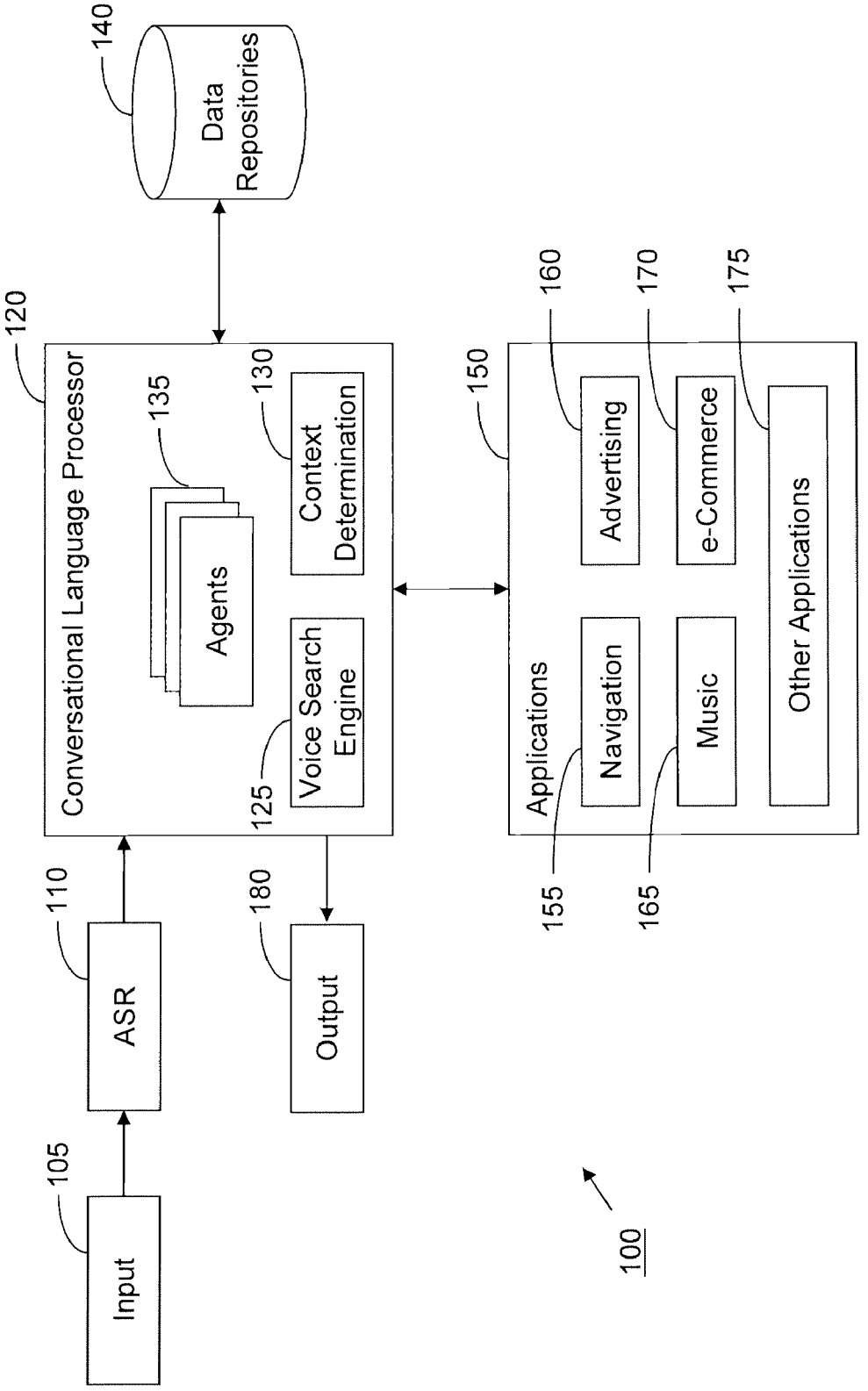

140
Data Repositories
120
Conversational Language Processor
135
Agents
130
Context Determination
125
Voice Search Engine
150
Applications
160
Advertising
170
e-Commerce
175
Other Applications
155
Navigation
165
Music
110
ASR
180
Output
105
Input
100

Fig. 1

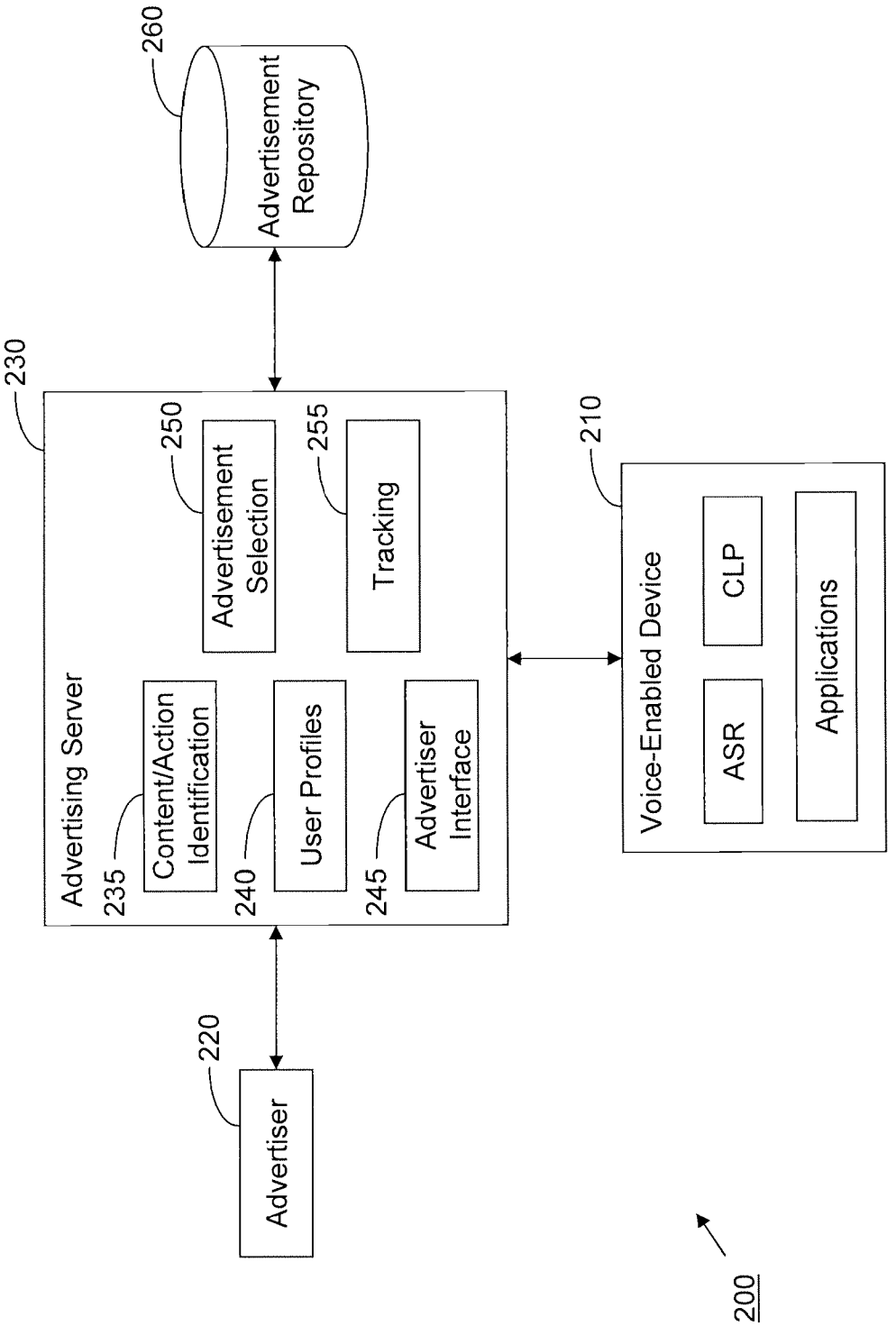

260
Advertisement Repository
230
250
Advertisement Selection
255
Tracking
210
Advertising Server
Content/Action Identification
235
User Profiles
240
Advertiser Interface
245
Voice-Enabled Device
CLP
ASR
Applications
220
Advertiser
200

Fig. 2

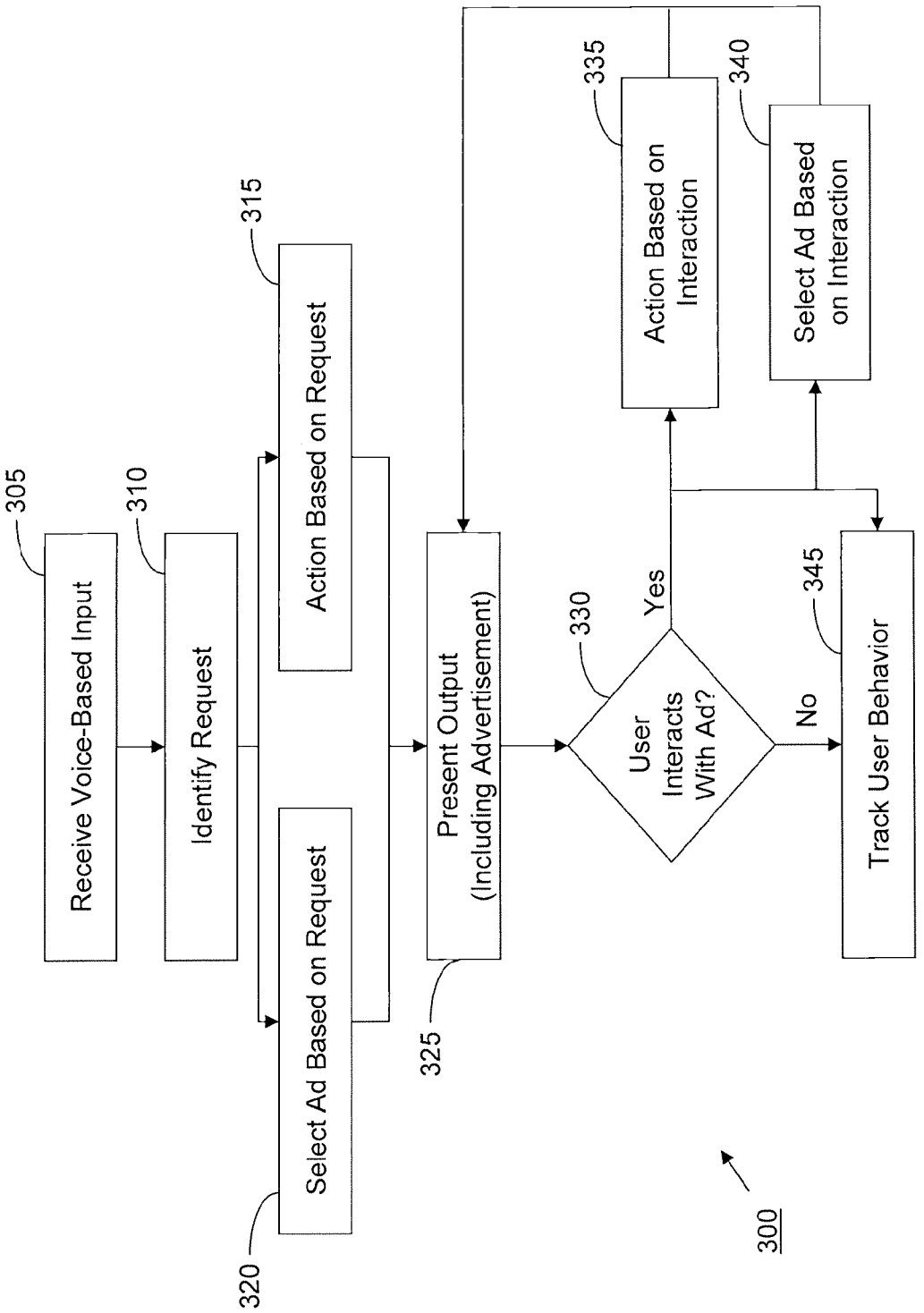

Receive Voice-Based Input
305
Identify Request
310
Action Based on Request
315
Select Ad Based on Request
320
Present Output (Including Advertisement)
325
User Interacts With Ad?
330
Yes
No
Action Based on Interaction
335
Select Ad Based on Interaction
340
Track User Behavior
345
300

Fig. 3

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

FIELD OF THE INVENTION

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

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

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. Patent 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 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 source. For example, shared knowledge sources may include information about previous utterances, requests, and other user interactions to inform generating an appropriate response to a current utterance. The shared knowledge may include public/non-private (i.e., environmental) knowledge, as well as personal/private (i.e., historical) knowledge. For example, conversational language processor **120** may use context determination module **130** to establish a context for a current utterance by having domain agents **135** competitively generate a context-based interpretation of the utterance (e.g., by scoring possible interpretations and selecting a highest scoring interpretation). As such, agents **135** may model various domains (e.g., navigation, music, a specific user, global users, advertising, e-commerce, etc.), and conversational language processor **120** may interpret and/or respond to a voice-based input accordingly. For example, context-based interpretations and responses to a voice-based input may be generated using techniques described in 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, 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, 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 envi-

ronmental 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 downloads, coupons, or any other suitable marketing, advertising, campaign, or other information, as would be apparent to those skilled in the art. A voice-enabled device **210** may receive a voice-based input and establish communications with advertising server **230**. Subsequently, advertising server **230** may select one or more advertisements from among the advertisements stored in data repository **260**, and the selected advertisements may be provided to the voice-enabled device for presentation to a user.

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 about the voice-based input, where advertisement selection module **250** may select one or more advertisements relevant to the voice-based input based on information extracted using content/action identification module **235** and/or user profile module **240**.

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, 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 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 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 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 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.

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 selecting and presenting advertisements in response to processing natural language utterances, comprising:

receiving a natural language utterance containing at least one request at an input device;

recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or phrases in the natural language utterance includes:

mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;

interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

selecting an advertisement in the context established for the natural language utterance; and

presenting the selected advertisement via an output device coupled to the conversational language processor.

2. The method of claim **1**, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

3. The method of claim **2**, further comprising:

tracking an interaction pattern with the advertisement presented via the output device; and

updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

4. The method of claim **3**, wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

5. The method of claim **4**, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

6. The method of claim **3**, further comprising:

building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model;

receiving a subsequent natural language utterance at the input device; and

interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.

7. The method of claim **3**, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

8. The method of claim **7**, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

9. The method of claim **1**, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

10. The method of claim **1**, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

11. The method of claim **10**, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

12. The method of claim **1**, further comprising determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

13. The method of claim **12**, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

14. A method for selecting and presenting advertisements in response to processing natural language utterances, comprising:

receiving a natural language utterance containing at least one request at an input device;

recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device;

interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

selecting an advertisement in the context established for the natural language utterance;

presenting the selected advertisement via an output device coupled to the conversational language processor; and

determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

15. The method of claim **14**, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

16. The method of claim **14**, wherein recognizing the words or phrases in the natural language utterance includes:

mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.

17. The method of claim **14**, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

18. The method of claim **14**, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

19. The method of claim **18**, further comprising:

tracking an interaction pattern with the advertisement presented via the output device; and

updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

20. The method of claim **19**, wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

21. The method of claim **20**, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

22. The method of claim **19**, further comprising:

building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model;

receiving a subsequent natural language utterance at the input device; and

interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.

23. The method of claim **19**, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

24. The method of claim **23**, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

25. The method of claim **14**, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one

or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

26. The method of claim **25**, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

27. A system for selecting and presenting advertisements in response to processing natural language utterances, comprising:

an input device that receives a natural language utterance containing at least one request at an input device;

a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to:

map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;

a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to:

interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

select an advertisement in the context established for the natural language utterance; and

present the selected advertisement via an output device.

28. The system of claim **27**, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

29. The system of claim **28** wherein the conversational language processor is further configured to:

track an interaction pattern with the advertisement presented via the output device; and

update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

30. The system of claim **29**, wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

31. The system of claim **30**, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

32. The system of claim **29**, wherein the conversational language processor is further configured to:

build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model; and

interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.

33. The system of claim **29**, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

34. The system of claim **33**, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

35. The system of claim **27**, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

36. The system of claim **27** wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

37. The system of claim **36**, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

38. The system of claim **27**, further comprising an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

39. The system of claim **38**, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

40. A system for selecting and presenting advertisements in response to processing natural language utterances, comprising:

an input device that receives a natural language utterance containing at least one request at an input device;

a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance;

a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to:

interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

select an advertisement in the context established for the natural language utterance; and

present the selected advertisement via an output device; and

an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

41. The system of claim **40**, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

42. The system of claim **40**, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to:

map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.

43. The system of claim **40**, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

44. The system of claim **40**, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

45. The system of claim **44**, wherein the conversational language processor is further configured to:

track an interaction pattern with the advertisement presented via the output device; and

update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

46. The system of claim **45**, wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

47. The system of claim **46**, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

48. The system of claim **45**, wherein the conversational language processor is further configured to:

build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model; and

interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.

49. The system of claim **45**, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

50. The system of claim **49**, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

51. The system of claim **40**, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

52. The system of claim **51** wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

* * * * *

EXHIBIT J

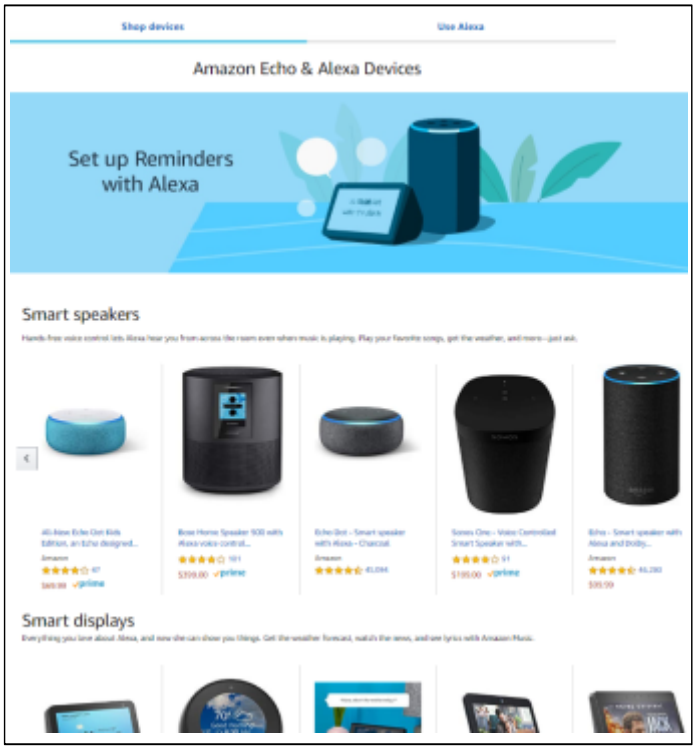
U.S. Pat. No. 7,818,176 ("the '176 patent")
Infringement Chart for Claim 27

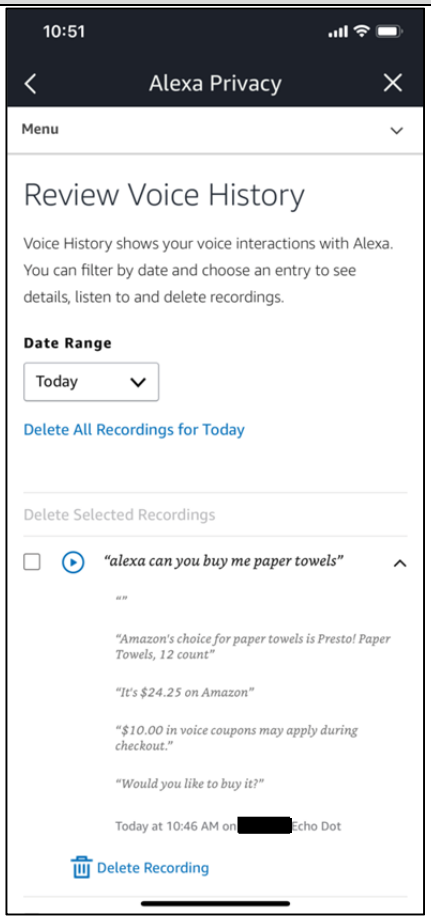
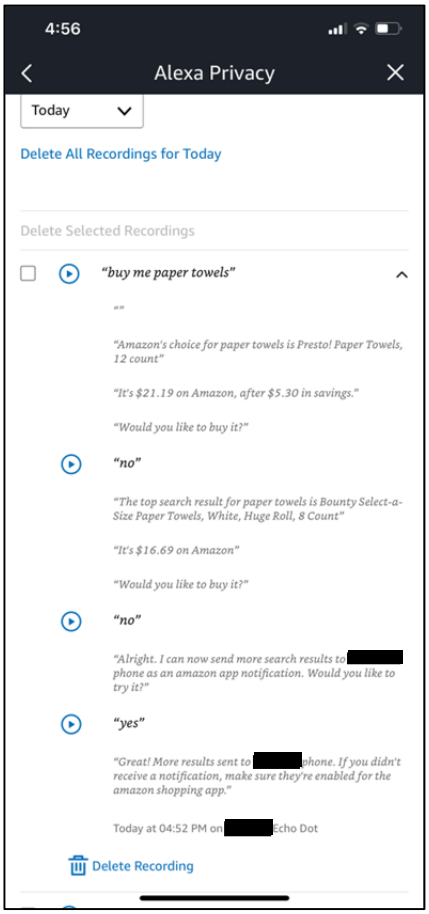
The following chart provides exemplary infringement allegations with respect to an exemplary claim of the '176 patent based on Plaintiff's investigation to date. Plaintiff reserves the right to provide modified and/or more detailed allegations (including with additional claims, products, and evidence) in accordance with its discovery obligations in view of any applicable scheduling order. This chart uses "Alexa Products" to refer to the full scope of instrumentalities identified in the complaint. Any discussion of specific Alexa Products in the following chart is exemplary.

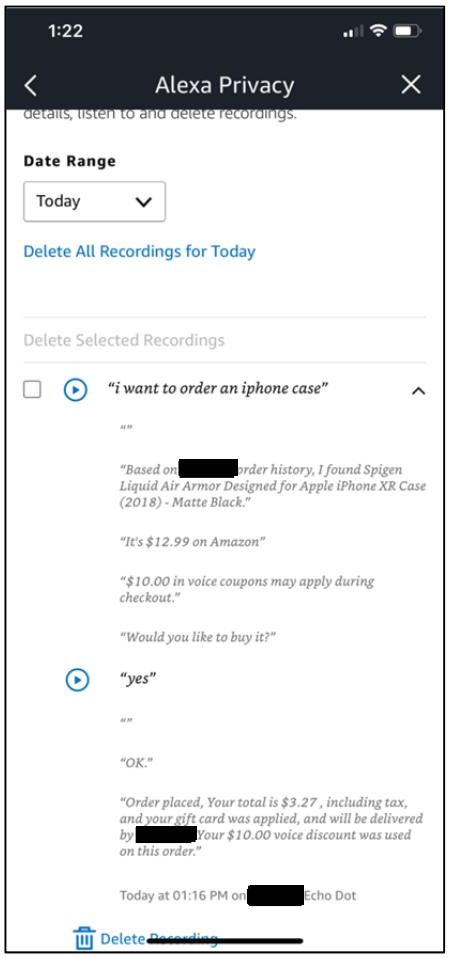
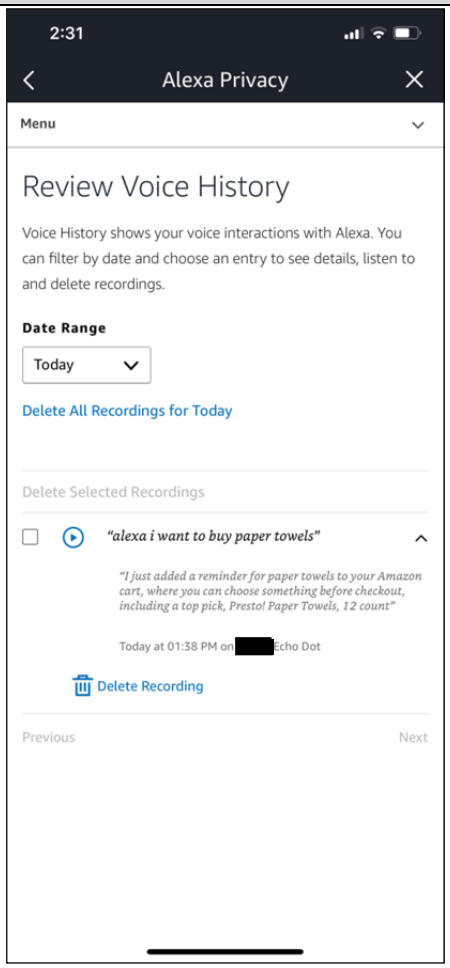
Exemplary information relevant to infringement:

1. *Amazon Echo & Alexa Devices* - Amazon Official Site, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/?ie=UTF8&node=9818047011&ref_=sv_devicesubnav_1 (last visited Jul. 22, 2019)
2. *Speech Recognizer Interface*, Alexa Voice Service, developer.amazon.com, https://developer.amazon.com/docs/alexa-voice-service/speechrecognizer.html (last visited Jul. 22, 2019)
3. Chaim Gartenberg, *How to shop on Amazon using Alexa voice controls*, The Verge, Jul. 10, 2017, https://www.theverge.com/2017/7/10/15947672/amazon-alexa-voice-controls-shopping-prime-echo-how-to ("*How to Shop*")
4. *Place Orders with Alexa*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201807210 (last visited Jul. 22, 2019)
5. *Alexa and Alexa Device FAQs*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (last visited Jul. 22, 2019)
6. *Audio Hardware Configurations*, Alexa Voice Service, developer.amazon.com, https://developer.amazon.com/docs/alexa-voice-service/audio-hardware-configurations.html (last visited Jul. 22, 2019)
7. Taylor Martin, *How to shop with Alexa on Prime Day*, CNET, Jul. 15, 2018, https://www.cnet.com/how-to/how-to-shop-with-alexa-prime-day/ ("*Prime Day*")
8. Richard Baguley & Colin McDonald, *Appliance Science: Alexa, how does Alexa work? The science of the Amazon Echo*, CNET, Aug. 4, 2016, https://www.cnet.com/news/appliance-science-alexa-how-does-alexa-work-the-science-of-amazons-echo/ ("*Appliance Science*")
9. *What is Conversational AI?*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/alexa-skills-kit/conversational-ai (last visited Jul. 22, 2019)
10. Sarah Perez, *Alexa will soon gain a memory, converse more naturally, and automatically launch skills,* TechCrunch, Apr. 26, 2018, https://techcrunch.com/2018/04/26/alexa-will-soon-gain-a-memory-converse-more-naturally-and-automatically-launch-skills/ ("*Alexa will soon gain a memory*")
11. *Designing for Conversation – Build for dialog*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/ask-resources/guided/conversational-design-workshop#/lessons/gDcQVF9nOFOtNljrELzARIGN4HuD7LNh?_k=q4qp1a (last visited Jul. 22, 2019)

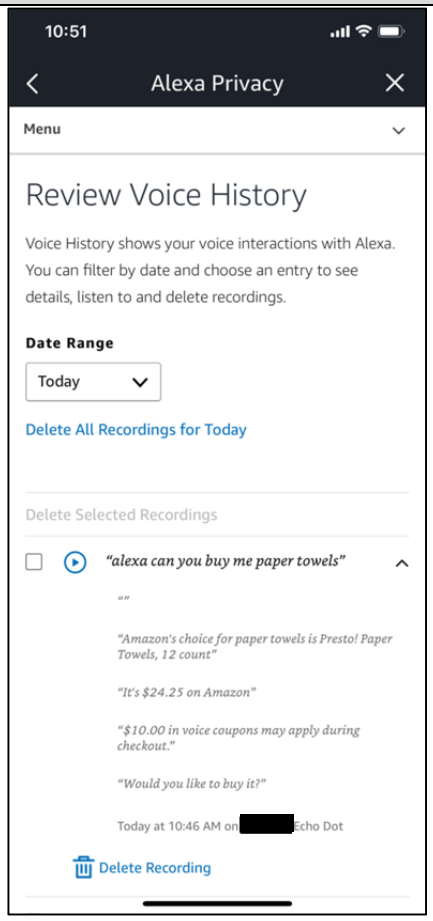
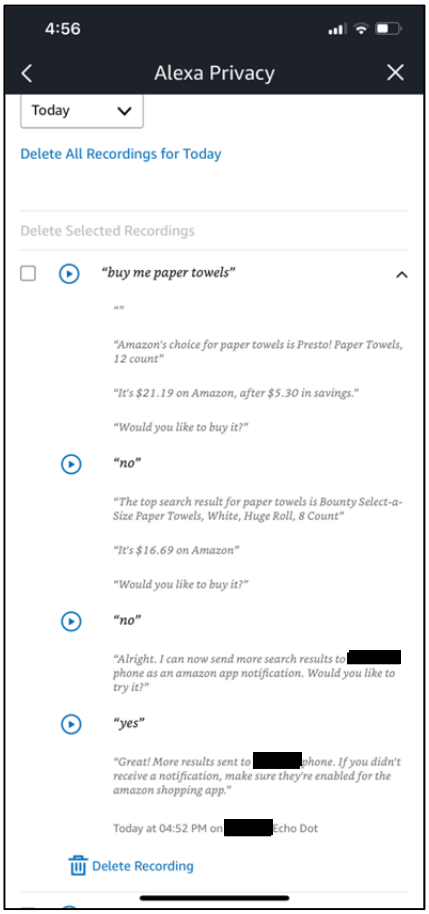
'176 Patent Claim Language	**Exemplary Alexa Products**
27.[pr] A system for selecting and presenting advertisements in response to processing natural	Alexa Products include a system for selecting and presenting advertisements in response to processing natural language utterances. For example, Alexa Products (e.g. Echo, Echo Dot, Alexa apps, and

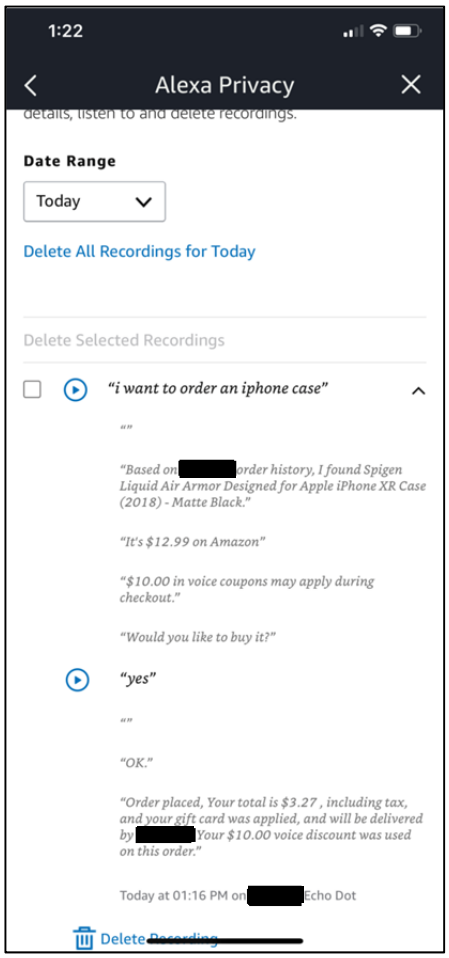
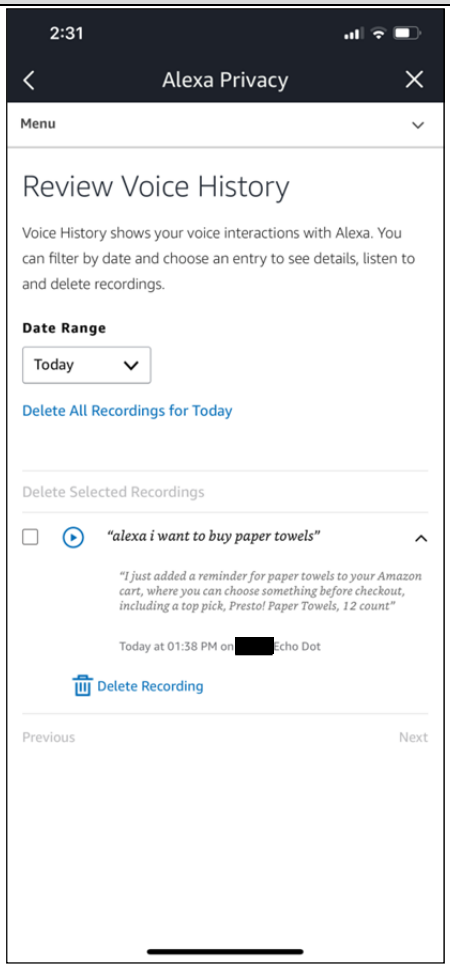
'176 Patent Claim Language	Exemplary Alexa Products
language utterances, comprising:	Alexa cloud) include Alexa.  E.g. *Amazon Echo & Alexa Devices*. "With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?* Alexa Products include Amazon's Alexa cloud. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1. When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*. "You can ask Alexa to place orders for Prime-eligible physical products. On supported devices, you can also ask Alexa to place orders for music" *Place Orders with Alexa*.
[27.a] an input device that receives a natural language utterance containing at least one request at an input device;	Alexa Products have an input device that receives a natural language utterance containing at least one request at an input device. Examples of receiving an utterance are shown in screenshots of Alexa app's "Review Voice History" page below.

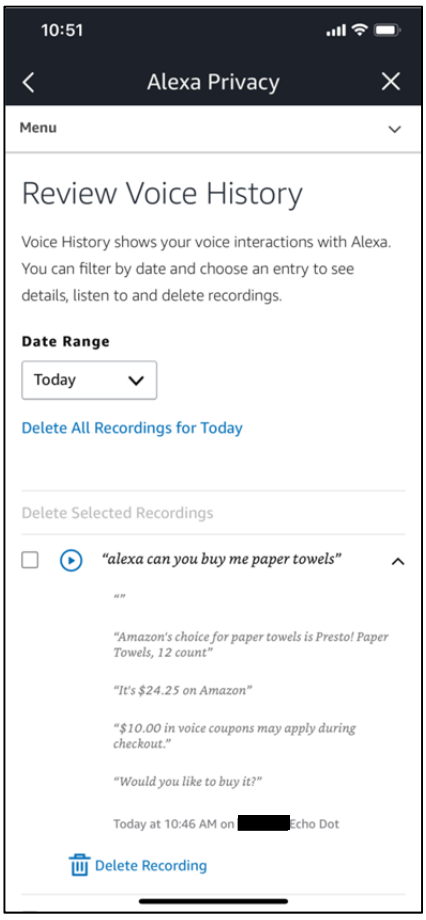
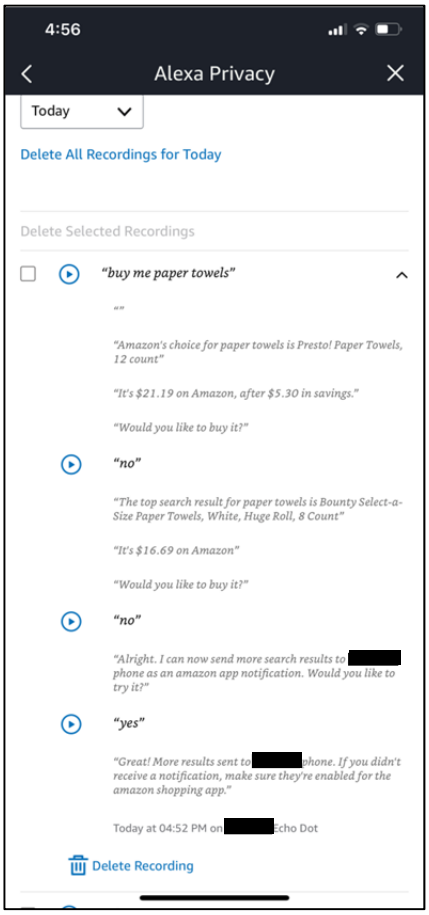
’176 Patent Claim Language	Exemplary Alexa Products
	 

'176 Patent Claim Language	Exemplary Alexa Products
	  Below is another example. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1.
[27.b] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to:	Alexa Products have a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance. For example: "When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask "Alexa, play top hits on Amazon Music" we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2. Another example is: "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product

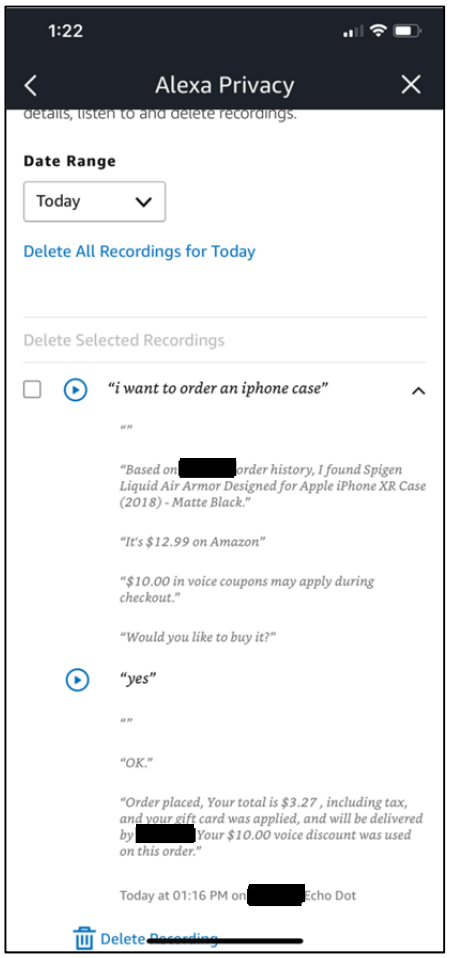
'176 Patent Claim Language	**Exemplary Alexa Products**
	instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles.
[27.c] map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and	Alexa Products map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar. For example: "When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask "Alexa, play top hits on Amazon Music" we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2. Another example is: "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles. Another example is: "Every user utterance leverages SpeechRecognizer. It is the core interface of the Alexa Voice Service (AVS). It exposes directives and events for capturing user speech and prompting a client when Alexa needs additional speech input." *Speech Recognizer Interface*
[27.d] generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;	Alexa Products generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases. Examples are generating preliminary interpretations for natural language utterances shown in screenshots of Alexa app's "Review Voice History" page below.

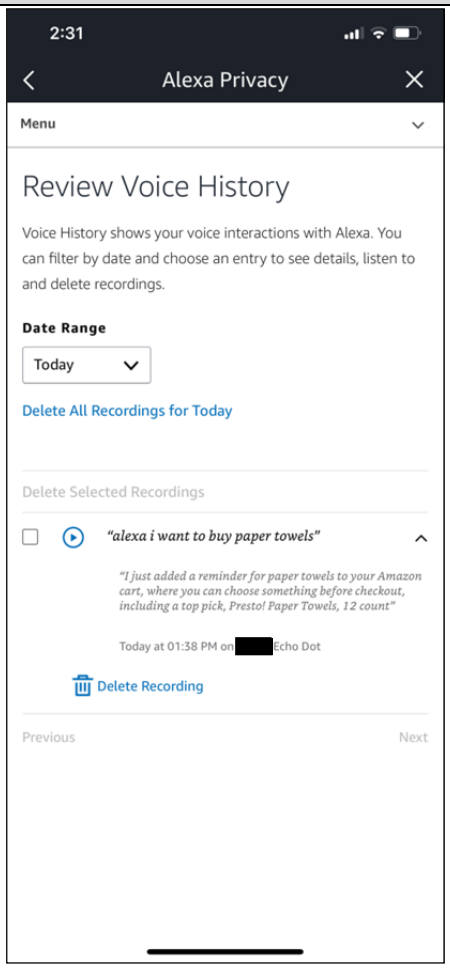
’176 Patent Claim Language	Exemplary Alexa Products
	 

'176 Patent Claim Language	Exemplary Alexa Products
	  Below is another example. "Every user utterance leverages SpeechRecognizer. It is the core interface of the Alexa Voice Service (AVS). It exposes directives and events for capturing user speech and prompting a client when Alexa needs additional speech input." *Speech Recognizer Interface*.
[27.e] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to:	Alexa Products have a conversational language processor coupled to the speech recognition engine. For example, Alexa Products include Amazon's Alexa cloud that has one or more processors: "When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you. For example, when you ask "Alexa, play top hits on Amazon Music" we use the recording of your request and information from Amazon Music to play top hits." *Alexa and Alexa Device FAQs* at answer to question 2. More examples are below. "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech

'176 Patent Claim Language	Exemplary Alexa Products
	Recognition Profiles. "The service that processes this recording is called Alexa Voice Services (AVS). Run by Amazon, this converts the recording into commands that it interprets. It's more than a simple voice-to-text service: it is a fully programmable service that can work with other online services to do a surprising range of things." *Appliance Science*.
[27.f] interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;	Alexa Products interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance. Examples are interpreting the recognized words or phrases shown in screenshots of Alexa app's "Review Voice History" page below.  

'176 Patent Claim Language	Exemplary Alexa Products





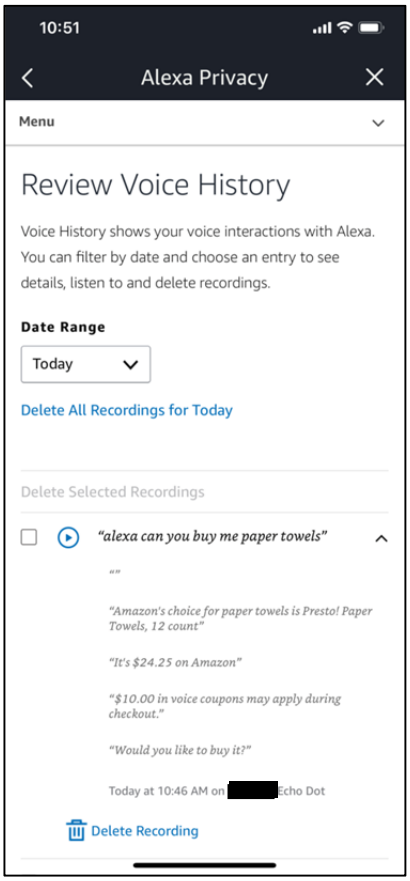
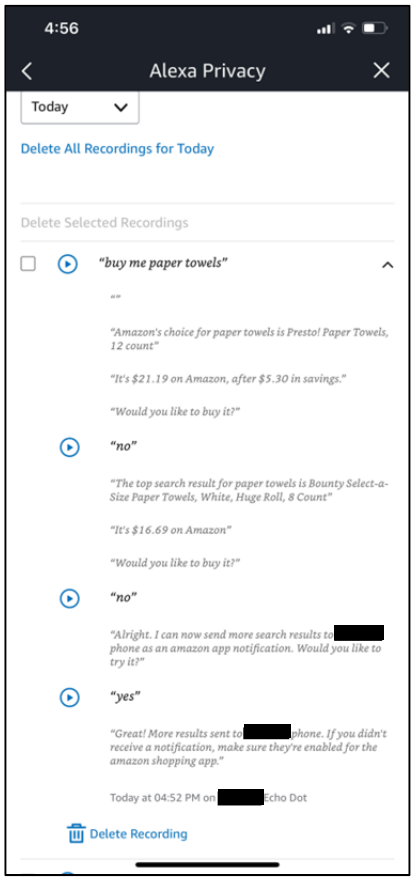
Below are more examples.

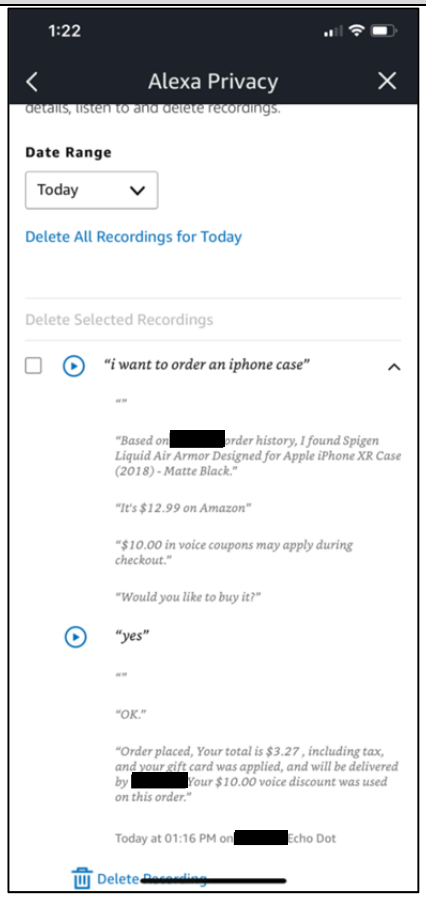
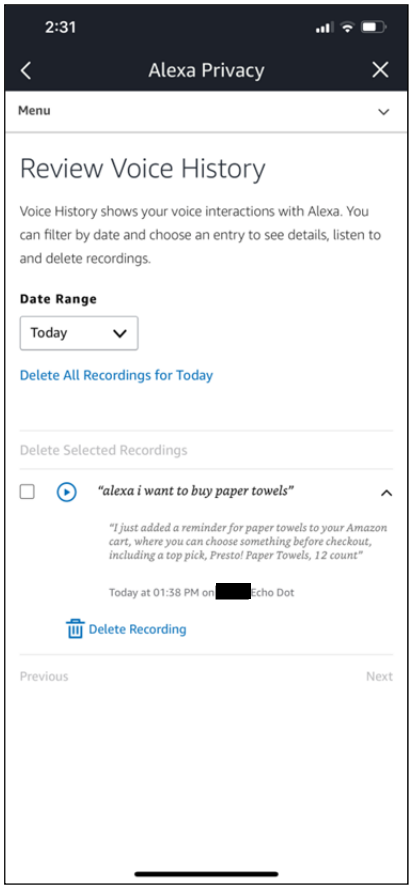
"Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles.

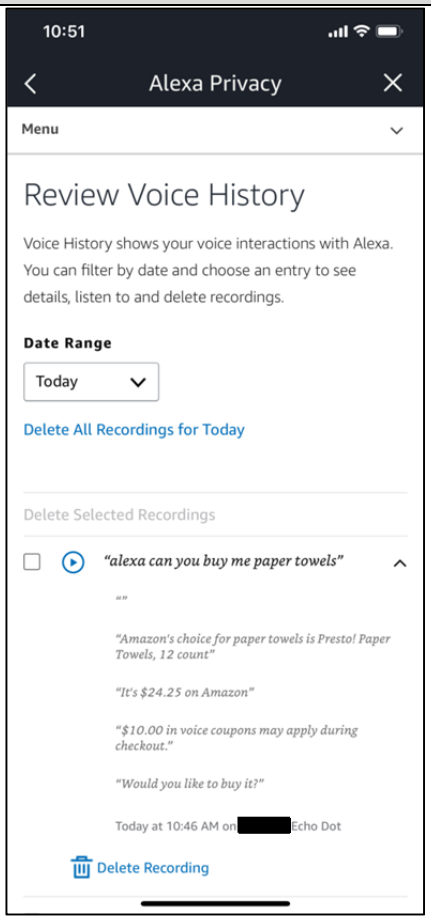
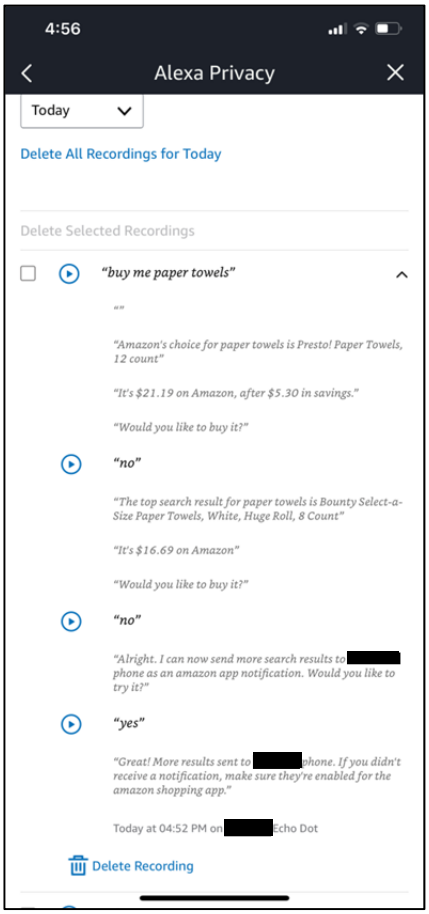
"The service that processes this recording is called Alexa Voice Services (AVS). Run by Amazon, this converts the recording into commands that it interprets. It's more than a simple voice-to-text service: it is a fully programmable service that can work with other online services to do a surprising range of things." *Appliance Science*.

"Everyday conversations take many things into account besides just the words we speak, including why, when, and where we converse. When designing for voice, think about how you can provide context to provide a delightful conversation." *What is Conversational AI?*

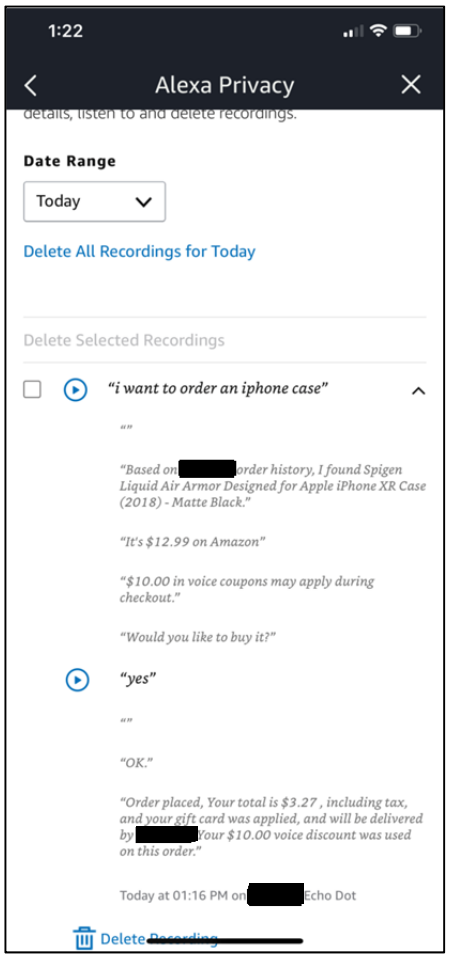
"Alexa will soon be able to have more natural conversations with users, thanks to something called 'context carryover.' This means that Alexa will be able to understand follow-up questions and respond appropriately..." *Alexa will soon gain a memory*.

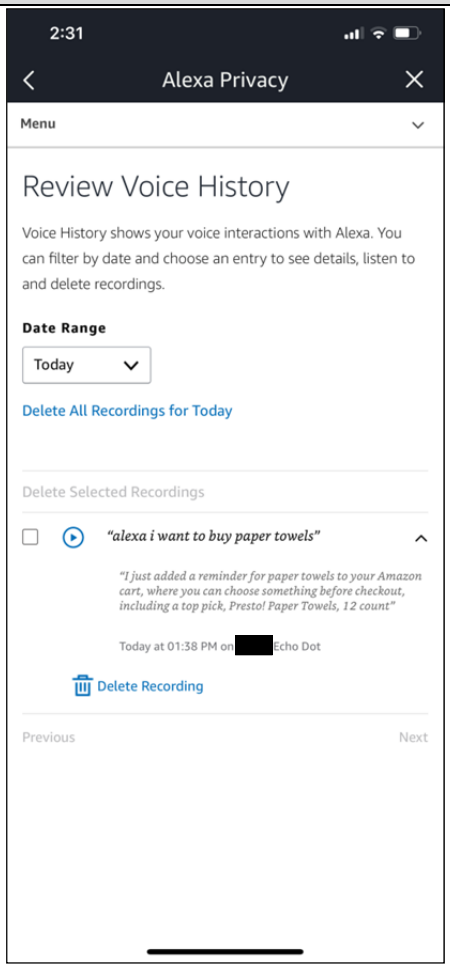
’176 Patent Claim Language	Exemplary Alexa Products
	“The person asking the question initiated the conversation, and you, having tracked the current state, pulled out the appropriate “file” to respond with an answer or appropriate followup question. Dialog Management provides a means for keeping track of context and prompting for required information regardless of the order in which the user provides it.” *Designing for Conversation – Build for dialog.*
[27.g] select an advertisement in the context established for the natural language utterance; and	Alexa Products select an advertisement in the context established for the natural language utterance. Examples are selecting advertisements as shown in screenshots of Alexa app’s “Review Voice History” page below.  

’176 Patent Claim Language	Exemplary Alexa Products
	  Below are more examples. “For starters, if you want to reorder something you've already purchased from Amazon before, just say, "Alexa, reorder a Swiss Army knife." Alexa will pull up your order history and let you select from the items. You confirm with Alexa which item you want from your history and you're done.” *Prime Day*. “To order something you've not ordered through Amazon before, say something like "Alexa, order deodorant," or "Alexa, buy a microSD card." Alexa will then perform a search for that item and offer a few recommendations based on Amazon preferred products or products you've purchased from similar categories in the past. It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add it to your cart.” *Prime Day*.
[27.h] present the selected advertisement via an output device.	Alexa Products present the selected advertisement via an output device Examples of presenting advertisements via an output device are shown in screenshots of Alexa app’s “Review Voice History” page below.

’176 Patent Claim Language	Exemplary Alexa Products
	 

’176 Patent Claim Language	Exemplary Alexa Products





Below are more examples.

“For starters, if you want to reorder something you've already purchased from Amazon before, just say, "Alexa, reorder a Swiss Army knife." Alexa will pull up your order history and let you select from the items. You confirm with Alexa which item you want from your history and you're done.” *Prime Day*.

“To order something you've not ordered through Amazon before, say something like "Alexa, order deodorant," or "Alexa, buy a microSD card." Alexa will then perform a search for that item and offer a few recommendations based on Amazon preferred products or products you've purchased from similar categories in the past. It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add it to your cart.” *Prime Day*.

EXHIBIT K

US008886536B2

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

(10) **Patent No.: US 8,886,536 B2**
(45) **Date of Patent: Nov. 11, 2014**

(54) **SYSTEM AND METHOD FOR DELIVERING TARGETED ADVERTISEMENTS AND TRACKING ADVERTISEMENT INTERACTIONS IN VOICE RECOGNITION CONTEXTS**

(71) Applicant: **VoiceBox Technologies Corporation**, Bellevue, WA (US)

(72) Inventors: **Tom Freeman**, Mercer Island, WA (US); **Mike Kennwick**, Bellevue, 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 0 days.

(21) Appl. No.: **14/016,757**

(22) Filed: **Sep. 3, 2013**

(65) **Prior Publication Data**

US 2014/0012577 A1 Jan. 9, 2014

Related U.S. Application Data

(63) 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.**

G10L 15/18	(2013.01)
G06F 17/27	(2006.01)
G10L 21/00	(2013.01)
G06Q 30/00	(2012.01)
G10L 15/26	(2006.01)
G06Q 30/02	(2012.01)

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

CPC ***G10L 15/18*** (2013.01); ***G10L 15/265*** (2013.01); ***G06Q 30/0251*** (2013.01); ***G06Q 30/0241*** (2013.01); ***G06Q 30/0242*** (2013.01)

USPC **704/257**; 704/9; 704/270.1; 705/14.4

(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
4,821,027	A	4/1989	Mallory et al.	340/521

(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)

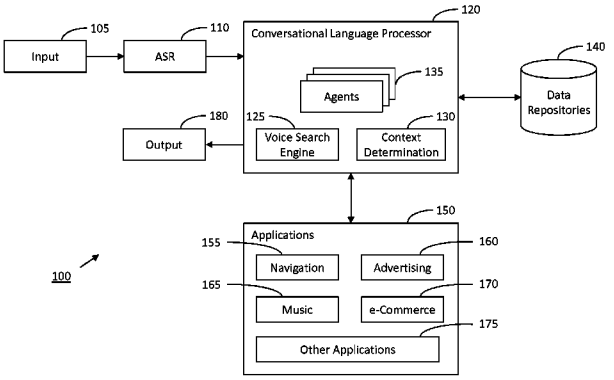
Primary Examiner — Brian Albertalli

(74) *Attorney, Agent, or Firm* — Pillsbury Winthrop Shaw Pittman LLP

(57) **ABSTRACT**

The system and method described herein may use various natural language models to deliver targeted advertisements and track advertisement interactions in voice recognition contexts. In particular, in response to an input device receiving an utterance, a conversational language processor may select and deliver one or more advertisements targeted to a user that spoke the utterance based on cognitive models associated with the user, various users having similar characteristics to the user, an environment in which the user spoke the utterance, or other criteria. Further, subsequent interaction with the targeted advertisements may be tracked to build and refine the cognitive models and thereby enhance the information used to deliver targeted advertisements in response to subsequent utterances.

55 Claims, 3 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

4,829,423	A	5/1989	Tennant et al.	364/200
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
5,274,560	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
5,475,733	A	12/1995	Eisdorfer et al.	379/52
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
5,617,407	A	4/1997	Bareis	369/275.3
5,633,922	A	5/1997	August et al.	379/220
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
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 et al.	395/2.66
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
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
5,832,221	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
5,867,817	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
5,902,347	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 et al.	707/1
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
5,960,397	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
5,987,404	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
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
6,076,059	A	6/2000	Glickman et al.	704/260
6,078,886	A	6/2000	Dragosh et al.	704/270
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
6,122,613	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
6,144,938	A	11/2000	Surace et al.	704/257
6,154,526	A	11/2000	Dahlke et al.	379/88.03
6,160,883	A	12/2000	Jackson et al.	379/230
6,167,377	A	12/2000	Gillick et al.	704/240
6,173,266	B1	1/2001	Marx et al.	704/270
6,173,279	B1	1/2001	Levin et al.	707/5
6,175,858	B1	1/2001	Bulfer et al.	709/206
6,185,535	B1	2/2001	Hedin et al.	704/270
6,188,982	B1	2/2001	Chiang	704/256
6,192,110	B1	2/2001	Abella et al.	379/88.01
6,192,338	B1	2/2001	Haszto et al.	704/257
6,195,634	B1	2/2001	Dudemaine et al.	704/231
6,195,651	B1	2/2001	Handel et al.	707/2
6,199,043	B1	3/2001	Happ	704/272
6,208,964	B1	3/2001	Sabourin	704/244
6,208,972	B1	3/2001	Grant et al.	704/275
6,219,346	B1	4/2001	Maxemchuk	370/338
6,219,643	B1	4/2001	Cohen et al.	704/257
6,226,612	B1	5/2001	Srenger et al.	704/256
6,233,556	B1	5/2001	Teunen et al.	704/250
6,233,559	B1	5/2001	Balakrishnan	704/275
6,233,561	B1	5/2001	Junqua et al.	704/277
6,236,968	B1	5/2001	Kanevsky et al.	704/275
6,246,981	B1	6/2001	Papineni et al.	704/235
6,246,990	B1	6/2001	Happ	704/275
6,266,636	B1	7/2001	Kosaka et al.	704/244
6,269,336	B1	7/2001	Ladd et al.	704/270
6,272,455	B1	8/2001	Hoshen et al.	704/1
6,275,231	B1	8/2001	Obradovich	345/349
6,278,377	B1	8/2001	DeLine et al.	340/815.4
6,278,968	B1	8/2001	Franz et al.	704/3
6,288,319	B1	9/2001	Catona	84/609
6,292,767	B1	9/2001	Jackson et al.	704/1
6,301,560	B1	10/2001	Masters	704/251
6,308,151	B1	10/2001	Smith	704/235
6,314,402	B1	11/2001	Monaco et al.	704/275
6,321,196	B1	11/2001	Franceschi	704/243
6,356,869	B1	3/2002	Chapados et al.	704/275
6,362,748	B1	3/2002	Huang	340/901
6,366,882	B1	4/2002	Bijl et al.	704/235
6,366,886	B1	4/2002	Dragosh et al.	704/270.1
6,374,214	B1	4/2002	Friedland et al.	704/235
6,377,913	B1	4/2002	Coffman et al.	704/8
6,381,535	B1	4/2002	Durocher et al.	701/202
6,385,596	B1	5/2002	Wiser et al.	705/51
6,385,646	B1	5/2002	Brown et al.	709/217
6,393,403	B1	5/2002	Majaniemi	704/275
6,393,428	B1	5/2002	Miller et al.	707/102
6,397,181	B1	5/2002	Li et al.	704/256
6,404,878	B1	6/2002	Jackson et al.	379/221.01
6,405,170	B1	6/2002	Phillips et al.	704/270
6,408,272	B1	6/2002	White et al.	704/270.1
6,411,810	B1	6/2002	Maxemchuk	455/453
6,411,893	B2	6/2002	Ruhl	701/207
6,415,257	B1	7/2002	Junqua et al.	704/275
6,418,210	B1	7/2002	Sayko	379/142.15

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,420,975	B1	7/2002	DeLine et al.	340/815.4
6,429,813	B2	8/2002	Feigen	342/357.13
6,430,285	B1	8/2002	Bauer et al.	379/265.01
6,430,531	B1	8/2002	Polish	704/257
6,434,523	B1	8/2002	Monaco	704/257
6,434,524	B1	8/2002	Weber	704/257
6,434,529	B1	8/2002	Walker et al.	704/275
6,442,522	B1	8/2002	Carberry et al.	704/257
6,446,114	B1	9/2002	Bulfer et al.	709/206
6,453,153	B1	9/2002	Bowker et al.	455/67.4
6,453,292	B2	9/2002	Ramaswamy et al.	704/235
6,456,711	B1	9/2002	Cheung et al.	379/265.09
6,456,974	B1	9/2002	Baker et al.	704/270.1
6,466,654	B1	10/2002	Cooper et al.	379/88.01
6,466,899	B1	10/2002	Yano et al.	704/1
6,470,315	B1	10/2002	Netsch et al.	704/256
6,487,494	B2	11/2002	Odinak et al.	701/202
6,487,495	B1	11/2002	Gale et al.	701/209
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
6,505,155	B1	1/2003	Vanbuskirk et al.	704/246
6,510,417	B1	1/2003	Woods et al.	704/275
6,513,006	B2	1/2003	Howard et al.	704/257
6,522,746	B1	2/2003	Marchok et al.	379/406.03
6,523,061	B1	2/2003	Halverson et al.	709/202
6,532,444	B1	3/2003	Weber	704/257
6,539,348	B1	3/2003	Bond et al.	704/9
6,549,629	B2	4/2003	Finn et al.	381/92
6,553,372	B1	4/2003	Brassell et al.	707/5
6,556,970	B1	4/2003	Sasaki et al.	704/257
6,556,973	B1	4/2003	Lewin	704/277
6,560,576	B1	5/2003	Cohen et al.	704/270
6,560,590	B1	5/2003	Shwe et al.	706/55
6,567,778	B1	5/2003	Chao Chang et al.	704/257
6,567,797	B1	5/2003	Schuetze et al.	707/2
6,570,555	B1	5/2003	Prevost et al.	345/156
6,570,964	B1	5/2003	Murveit et al.	379/67.1
6,571,279	B1	5/2003	Herz et al.	709/217
6,574,597	B1	6/2003	Mohri et al.	704/251
6,574,624	B1	6/2003	Johnson et al.	707/5
6,578,022	B1	6/2003	Foulger et al.	706/45
6,581,103	B1	6/2003	Dengler	709/231
6,584,439	B1	6/2003	Geilhufe et al.	704/270
6,587,858	B1	7/2003	Strazza	707/102
6,591,239	B1	7/2003	McCall et al.	704/275
6,594,257	B1	7/2003	Doshi et al.	370/352
6,594,367	B1	7/2003	Marash et al.	381/92
6,598,018	B1	7/2003	Junqua	704/251
6,601,026	B2	7/2003	Appelt et al.	704/9
6,604,075	B1	8/2003	Brown et al.	704/270.1
6,604,077	B2	8/2003	Dragosh et al.	704/270.1
6,606,598	B1	8/2003	Holthouse et al.	704/275
6,611,692	B2	8/2003	Raffel et al.	455/552
6,614,773	B1	9/2003	Maxemchuk	370/337
6,615,172	B1	9/2003	Bennett et al.	704/257
6,622,119	B1	9/2003	Ramaswamy et al.	704/9
6,629,066	B1	9/2003	Jackson et al.	704/9
6,631,346	B1	10/2003	Karaorman et al.	704/9
6,631,351	B1	10/2003	Ramachandran et al.	704/270
6,633,846	B1	10/2003	Bennett et al.	704/257
6,636,790	B1	10/2003	Lightner et al.	701/33
6,643,620	B1	11/2003	Contolini et al.	704/270
6,650,747	B1	11/2003	Bala et al.	379/265.06
6,658,388	B1	12/2003	Kleindienst et al.	704/275
6,678,680	B1	1/2004	Woo	707/6
6,681,206	B1	1/2004	Gorin et al.	704/243
6,691,151	B1	2/2004	Cheyer 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
6,721,633	B2	4/2004	Funk et al.	701/1
6,721,706	B1	4/2004	Strubbe et al.	704/275
6,726,636	B2	4/2004	Der Ghazarian et al.	600/532
6,735,592	B1	5/2004	Neumann et al.	707/101
6,739,556	B1	5/2004	Langston	244/189
6,741,931	B1	5/2004	Kohut et al.	701/209
6,742,021	B1	5/2004	Halverson et al.	709/218
6,745,161	B1	6/2004	Arnold et al.	704/7
6,751,591	B1	6/2004	Gorin et al.	704/257
6,751,612	B1	6/2004	Schuetze et al.	707/4
6,754,485	B1	6/2004	Obradovich et al.	455/414.1
6,754,627	B2	6/2004	Woodward	704/235
6,757,544	B2	6/2004	Rangarajan et al.	455/456.1
6,757,718	B1	6/2004	Halverson et al.	709/218
6,795,808	B1	9/2004	Strubbe et al.	704/275
6,801,604	B2	10/2004	Maes et al.	379/88.17
6,801,893	B1	10/2004	Backfried et al.	704/257
6,813,341	B1	11/2004	Mahoney	379/88.01
6,829,603	B1	12/2004	Chai et al.	707/5
6,832,230	B1	12/2004	Zilliacus et al.	707/203
6,833,848	B1	12/2004	Wolff et al.	345/719
6,850,603	B1	2/2005	Eberle et al.	379/88.16
6,856,990	B2	2/2005	Barile et al.	707/10
6,865,481	B2	3/2005	Kawazoe et al.	701/211
6,868,380	B2	3/2005	Kroeker	704/240
6,868,385	B1	3/2005	Gerson	704/275
6,873,837	B1	3/2005	Yoshioka et al.	455/321
6,877,001	B2	4/2005	Wolf et al.	707/3
6,877,134	B1	4/2005	Fuller et al.	715/500.1
6,901,366	B1	5/2005	Kuhn et al.	704/275
6,910,003	B1	6/2005	Arnold et al.	704/4
6,912,498	B2	6/2005	Stevens et al.	704/235
6,915,126	B2	7/2005	Mazzara, Jr.	455/411
6,928,614	B1	8/2005	Everhart	715/728
6,934,756	B2	8/2005	Maes	709/227
6,937,977	B2	8/2005	Gerson	704/201
6,937,982	B2	8/2005	Kitaoka et al.	704/252
6,944,594	B2	9/2005	Busayapongchai et al.	704/275
6,950,821	B2	9/2005	Faybishenko et al.	707/10
6,954,755	B2	10/2005	Reisman	707/10
6,959,276	B2	10/2005	Droppo et al.	704/226
6,961,700	B2	11/2005	Mitchell et al.	704/235
6,963,759	B1	11/2005	Gerson	455/563
6,964,023	B2	11/2005	Maes et al.	715/811
6,968,311	B2	11/2005	Knockeart et al.	704/270
6,973,387	B2	12/2005	Masclet et al.	701/211
6,975,993	B1	12/2005	Keiller	704/275
6,980,092	B2	12/2005	Turnbull et al.	340/425.5
6,983,055	B2	1/2006	Luo	381/313
6,990,513	B2	1/2006	Belfiore et al.	709/203
6,996,531	B2	2/2006	Korall et al.	704/270
7,003,463	B1	2/2006	Maes et al.	704/270.1
7,016,849	B2	3/2006	Arnold et al.	704/275
7,020,609	B2	3/2006	Thrift et al.	704/270.1
7,024,364	B2	4/2006	Guerra et al.	704/270
7,027,586	B2	4/2006	Bushey et al.	379/265.09
7,027,975	B1	4/2006	Pazandak et al.	704/9
7,035,415	B2	4/2006	Belt et al.	381/92
7,036,128	B1	4/2006	Julia et al.	719/317
7,043,425	B2	5/2006	Pao	704/211
7,054,817	B2	5/2006	Shao	704/270
7,058,890	B2	6/2006	George et al.	715/728
7,062,488	B1	6/2006	Reisman	707/8
7,069,220	B2	6/2006	Coffman et al.	704/275
7,072,834	B2	7/2006	Zhou	704/244
7,076,362	B2	7/2006	Ohtsuji et al.	701/211
7,082,469	B2	7/2006	Gold et al.	709/231
7,085,708	B2	8/2006	Manson	704/9
7,092,928	B1	8/2006	Elad et al.	706/60
7,107,210	B2	9/2006	Deng et al.	704/226
7,107,218	B1	9/2006	Preston	704/270
7,110,951	B1	9/2006	Lemelson et al.	704/270
7,127,400	B2	10/2006	Koch	704/270.1
7,130,390	B2	10/2006	Abburi	379/88.17
7,136,875	B2	11/2006	Anderson et al.	707/104.1
7,137,126	B1	11/2006	Coffman et al.	719/328
7,143,037	B1	11/2006	Chestnut	704/251
7,143,039	B1	11/2006	Stifelman et al.	704/270

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,146,319	B2	12/2006	Hunt	704/254
7,149,696	B2	12/2006	Shimizu et al.	705/1
7,165,028	B2	1/2007	Gong	704/233
7,170,993	B2 *	1/2007	Anderson et al.	379/265.09
7,171,291	B2	1/2007	Obradovich	701/29
7,174,300	B2	2/2007	Bush	704/275
7,177,798	B2	2/2007	Hsu et al.	704/9
7,184,957	B2	2/2007	Brookes et al.	704/246
7,190,770	B2	3/2007	Ando et al.	379/88.01
7,197,069	B2	3/2007	Agazzi et al.	375/233
7,197,460	B1	3/2007	Gupta et al.	704/270.1
7,203,644	B2	4/2007	Anderson et al.	704/246
7,206,418	B2	4/2007	Yang et al.	381/92
7,207,011	B2	4/2007	Mulvey et al.	715/812
7,215,941	B2	5/2007	Beckmann et al.	455/404.1
7,228,276	B2	6/2007	Omote et al.	704/243
7,231,343	B1	6/2007	Treadgold et al.	704/9
7,236,923	B1	6/2007	Gupta	704/9
7,254,482	B2	8/2007	Kawasaki et al.	701/211
7,272,212	B2	9/2007	Eberle et al.	379/88.17
7,277,854	B2	10/2007	Bennett et al.	704/257
7,283,829	B2	10/2007	Christenson et al.	455/461
7,283,951	B2	10/2007	Marchisio et al.	704/9
7,289,606	B2	10/2007	Sibal et al.	379/52
7,299,186	B2	11/2007	Kuzunuki et al.	704/270.1
7,301,093	B2	11/2007	Sater et al.	84/615
7,305,381	B1	12/2007	Poppink et al.	1/1
7,321,850	B2	1/2008	Wakita	704/10
7,328,155	B2	2/2008	Endo et al.	704/251
7,337,116	B2	2/2008	Charlesworth et al.	704/254
7,340,040	B1	3/2008	Saylor et al.	379/67.1
7,366,285	B2	4/2008	Parolkar et al.	379/88.17
7,366,669	B2	4/2008	Nishitani et al.	704/256
7,376,645	B2	5/2008	Bernard	707/3
7,386,443	B1	6/2008	Parthasarathy et al.	704/201
7,398,209	B2	7/2008	Kennewick et al.	704/255
7,406,421	B2	7/2008	Odinak et al.	704/275
7,415,414	B2	8/2008	Azara et al.	704/270
7,421,393	B1	9/2008	Di Fabbrizio et al.	704/275
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
7,454,608	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
7,472,075	B2	12/2008	Odinak et al.	705/26
7,477,909	B2 *	1/2009	Roth	455/466
7,478,036	B2	1/2009	Shen et al.	704/9
7,487,088	B1	2/2009	Gorin et al.	704/240
7,487,110	B2	2/2009	Bennett et al.	705/26
7,493,259	B2	2/2009	Jones et al.	704/257
7,493,559	B1	2/2009	Wolff et al.	715/727
7,502,672	B1	3/2009	Kolls	701/29
7,502,738	B2	3/2009	Kennewick et al.	704/257
7,516,076	B2	4/2009	Walker et al.	704/275
7,529,675	B2	5/2009	Maes	704/270.1
7,536,297	B2	5/2009	Byrd et al.	704/10
7,536,374	B2	5/2009	Au	706/55
7,542,894	B2	6/2009	Murata	704/9
7,546,382	B2	6/2009	Healey et al.	709/246
7,548,491	B2	6/2009	Macfarlane	367/198
7,552,054	B1 *	6/2009	Stifelman et al.	704/270
7,558,730	B2	7/2009	Davis et al.	704/235
7,574,362	B2	8/2009	Walker et al.	704/275
7,577,244	B2	8/2009	Taschereau	379/218.01
7,606,708	B2	10/2009	Hwang	704/257
7,620,549	B2	11/2009	Di Cristo et al.	704/257
7,634,409	B2	12/2009	Kennewick et al.	704/257
7,640,006	B2	12/2009	Portman et al.	455/412.1
7,640,160	B2	12/2009	Di Cristo et al.	704/257
7,640,272	B2	12/2009	Mahajan et al.	707/104.1
7,676,365	B2	3/2010	Hwang et al.	704/240
7,676,369	B2	3/2010	Fujimoto et al.	704/270
7,684,977	B2	3/2010	Morikawa	704/211
7,693,720	B2	4/2010	Kennewick et al.	704/275
7,729,916	B2	6/2010	Coffman et al.	704/270
7,729,918	B2	6/2010	Walker et al.	704/275
7,729,920	B2	6/2010	Chaar et al.	704/275
7,734,287	B2	6/2010	Ying	455/423
7,748,021	B2	6/2010	Obradovich	725/105
7,788,084	B2	8/2010	Brun et al.	704/7
7,801,731	B2	9/2010	Odinak et al.	704/275
7,809,570	B2	10/2010	Kennewick et al.	704/257
7,818,176	B2	10/2010	Freeman et al.	704/270
7,831,426	B2	11/2010	Bennett	704/252
7,831,433	B1	11/2010	Belvin et al.	704/275
7,856,358	B2 *	12/2010	Ho	704/270
7,873,519	B2	1/2011	Bennett	704/257
7,873,523	B2	1/2011	Potter et al.	704/275
7,873,654	B2	1/2011	Bernard	707/769
7,881,936	B2	2/2011	Longe et al.	704/257
7,890,324	B2	2/2011	Bangalore et al.	704/231
7,894,849	B2	2/2011	Kass et al.	455/550.1
7,902,969	B2	3/2011	Obradovich	340/439
7,917,367	B2	3/2011	Di Cristo et al.	704/270.1
7,920,682	B2	4/2011	Byrne et al.	379/88.18
7,949,529	B2	5/2011	Weider et al.	704/270
7,949,537	B2	5/2011	Walker et al.	704/275
7,953,732	B2	5/2011	Frank et al.	707/724
7,974,875	B1	7/2011	Quilici et al.	705/14.4
7,983,917	B2	7/2011	Kennewick et al.	704/257
7,984,287	B2	7/2011	Gopalakrishnan et al.	713/100
8,005,683	B2	8/2011	Tessel et al.	704/275
8,015,006	B2	9/2011	Kennewick et al.	704/236
8,060,367	B2	11/2011	Keaveney	704/247
8,069,046	B2	11/2011	Kennewick et al.	704/257
8,073,681	B2	12/2011	Baldwin et al.	704/9
8,077,975	B2	12/2011	Ma et al.	382/187
8,082,153	B2	12/2011	Coffman et al.	704/270
8,086,463	B2	12/2011	Ativanichayaphong et al.	704/275
8,112,275	B2	2/2012	Kennewick et al.	704/240
8,140,327	B2	3/2012	Kennewick et al.	704/226
8,140,335	B2	3/2012	Kennewick et al.	704/257
8,145,489	B2	3/2012	Freeman et al.	704/257
8,150,694	B2	4/2012	Kennewick et al.	704/257
8,155,962	B2	4/2012	Kennewick et al.	704/257
8,170,867	B2	5/2012	Germain	704/9
8,195,468	B2	6/2012	Weider et al.	704/275
8,219,399	B2	7/2012	Lutz et al.	704/270
8,219,599	B2	7/2012	Tunstall-Pedoe	707/955
8,224,652	B2	7/2012	Wang et al.	704/275
8,255,224	B2	8/2012	Singleton et al.	704/275
8,326,627	B2	12/2012	Kennewick et al.	704/257
8,326,634	B2	12/2012	Di Cristo et al.	704/270.1
8,326,637	B2	12/2012	Baldwin et al.	704/275
8,332,224	B2	12/2012	Di Cristo et al.	704/257
8,370,147	B2	2/2013	Kennewick et al.	704/257
8,447,607	B2	5/2013	Weider et al.	704/250
8,452,598	B2	5/2013	Kennewick et al.	704/257
8,509,403	B2 *	8/2013	Chiu et al.	379/114.13
8,515,765	B2	8/2013	Baldwin et al.	704/275
8,527,274	B2	9/2013	Freeman et al.	704/257
8,589,161	B2	11/2013	Kennewick et al.	704/252
8,620,659	B2	12/2013	Di Cristo et al.	704/257
8,719,009	B2	5/2014	Baldwin et al.	704/9
8,719,026	B2	5/2014	Kennewick et al.	704/257
8,738,380	B2	5/2014	Baldwin et al.	704/257
2001/0041980	A1	11/2001	Howard et al.	704/270
2001/0049601	A1	12/2001	Kroeker et al.	704/254
2001/0054087	A1	12/2001	Flom et al.	709/218
2002/0015500	A1	2/2002	Belt et al.	381/66
2002/0022927	A1	2/2002	Lemelson et al.	701/301
2002/0029261	A1	3/2002	Shibata	709/219
2002/0032752	A1	3/2002	Gold et al.	709/218
2002/0035501	A1	3/2002	Handel et al.	705/10
2002/0040297	A1	4/2002	Tsiao et al.	704/257
2002/0049535	A1	4/2002	Rigo et al.	701/211
2002/0049805	A1	4/2002	Yamada et al.	709/202
2002/0065568	A1	5/2002	Silfvast et al.	700/94
2002/0067839	A1	6/2002	Heinrich	381/110
2002/0069059	A1	6/2002	Smith	704/257
2002/0069071	A1	6/2002	Knockeart et al.	704/275

(56) **References Cited**

U.S. PATENT DOCUMENTS

2002/0082911	A1	6/2002	Dunn et al.	705/14
2002/0087326	A1	7/2002	Lee et al.	704/270.1
2002/0087525	A1	7/2002	Abbott et al.	707/3
2002/0107694	A1	8/2002	Lerg	704/273
2002/0120609	A1	8/2002	Lang et al.	707/1
2002/0124050	A1	9/2002	Middeljans	709/203
2002/0133354	A1	9/2002	Ross et al.	704/275
2002/0133402	A1	9/2002	Faber et al.	705/14
2002/0135618	A1	9/2002	Maes et al.	345/767
2002/0138248	A1	9/2002	Corston-Oliver et al.	704/1
2002/0143532	A1	10/2002	McLean et al.	704/235
2002/0143535	A1	10/2002	Kist et al.	704/251
2002/0161646	A1	10/2002	Gailey et al.	705/14
2002/0173333	A1	11/2002	Buchholz et al.	455/527
2002/0173961	A1	11/2002	Guerra	704/258
2002/0184373	A1	12/2002	Maes	709/228
2002/0188602	A1	12/2002	Stubler et al.	707/3
2002/0198714	A1	12/2002	Zhou	704/252
2003/0014261	A1	1/2003	Kageyama	704/275
2003/0016835	A1	1/2003	Elko et al.	381/92
2003/0046346	A1	3/2003	Mumick et al.	709/205
2003/0064709	A1	4/2003	Gailey et al.	455/412
2003/0065427	A1	4/2003	Funk et al.	701/1
2003/0069734	A1	4/2003	Everhart	704/275
2003/0088421	A1	5/2003	Maes et al.	704/270.1
2003/0097249	A1	5/2003	Walker et al.	704/1
2003/0110037	A1	6/2003	Walker et al.	704/257
2003/0112267	A1	6/2003	Belrose	345/728
2003/0115062	A1	6/2003	Walker et al.	704/258
2003/0120493	A1	6/2003	Gupta	704/270.1
2003/0135488	A1	7/2003	Amir et al.	707/3
2003/0144846	A1	7/2003	Denenberg et al.	704/277
2003/0158731	A1	8/2003	Falcon et al.	704/231
2003/0161448	A1	8/2003	Parolkar et al.	379/88.17
2003/0182132	A1	9/2003	Niemoeller	704/275
2003/0204492	A1	10/2003	Wolf et al.	707/3
2003/0206640	A1	11/2003	Malvar et al.	381/93
2003/0212550	A1	11/2003	Ubale	704/215
2003/0212558	A1	11/2003	Matula	704/260
2003/0212562	A1	11/2003	Patel et al.	704/275
2003/0225825	A1	12/2003	Healey et al.	709/203
2003/0236664	A1	12/2003	Sharma	704/251
2004/0006475	A1	1/2004	Ehlen et al.	704/270.1
2004/0010358	A1	1/2004	Oesterling et al.	701/49
2004/0025115	A1	2/2004	Sienel et al.	715/513
2004/0044516	A1	3/2004	Kennewick et al.	704/5
2004/0098245	A1	5/2004	Walker et al.	704/1
2004/0117179	A1	6/2004	Balasuriya	704/231
2004/0117804	A1	6/2004	Scahill et al.	719/320
2004/0122674	A1	6/2004	Bangalore et al.	704/276
2004/0140989	A1	7/2004	Papageorge	345/700
2004/0158555	A1	8/2004	Seedman et al.	707/3
2004/0166832	A1	8/2004	Portman et al.	455/412.1
2004/0167771	A1	8/2004	Duan et al.	704/10
2004/0172258	A1	9/2004	Dominach et al.	704/277
2004/0193408	A1	9/2004	Hunt	704/209
2004/0193420	A1	9/2004	Kennewick et al.	704/257
2004/0199375	A1	10/2004	Ehsani et al.	704/4
2004/0205671	A1	10/2004	Sukehiro et al.	715/532
2004/0243417	A9	12/2004	Pitts, III et al.	704/276
2005/0015256	A1	1/2005	Kargman	704/272
2005/0021334	A1	1/2005	Iwahashi	704/240
2005/0021470	A1	1/2005	Martin et al.	705/51
2005/0021826	A1	1/2005	Kumar	709/232
2005/0033574	A1	2/2005	Kim et al.	704/251
2005/0033582	A1*	2/2005	Gadd et al.	704/277
2005/0043940	A1	2/2005	Elder	704/9
2005/0114116	A1	5/2005	Fiedler	704/201
2005/0125232	A1	6/2005	Gadd	704/270.1
2005/0131673	A1	6/2005	Koizumi et al.	704/2
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
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
2007/0043569	A1	2/2007	Potter, III et al.	704/270
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
2007/0061067	A1	3/2007	Zeinstra et al.	701/200
2007/0061735	A1	3/2007	Hoffberg et al.	715/744
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
2007/0203736	A1*	8/2007	Ashton	705/1
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
2008/0147410	A1	6/2008	Odinak	704/270.1
2008/0154604	A1	6/2008	Sathish et al.	704/257
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
2009/0307031	A1	12/2009	Winkler et al.	705/7
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
2010/0145700	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
2011/0112827	A1	5/2011	Kennewick et al.	704/9

(56) **References Cited**

U.S. PATENT DOCUMENTS

2011/0112921	A1	5/2011	Kennewick et al.	705/26.1
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
2013/0054228	A1	2/2013	Baldwin et al.	704/9
2013/0211710	A1	8/2013	Kennewick et al.	701/419
2013/0253929	A1	9/2013	Weider et al.	704/235
2013/0297293	A1	11/2013	Di Cristo et al.	704/9
2013/0304473	A1	11/2013	Baldwin et al.	704/257
2013/0339022	A1	12/2013	Baldwin et al.	704/257
2014/0108013	A1	4/2014	Di Cristo et al.	704/254
2014/0156278	A1	6/2014	Kennewick et al.	704/254

FOREIGN PATENT DOCUMENTS

JP	2006-146881	6/2006
JP	2008-027454	2/2008
JP	2008-139928	6/2008
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.

Cheyer 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.

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

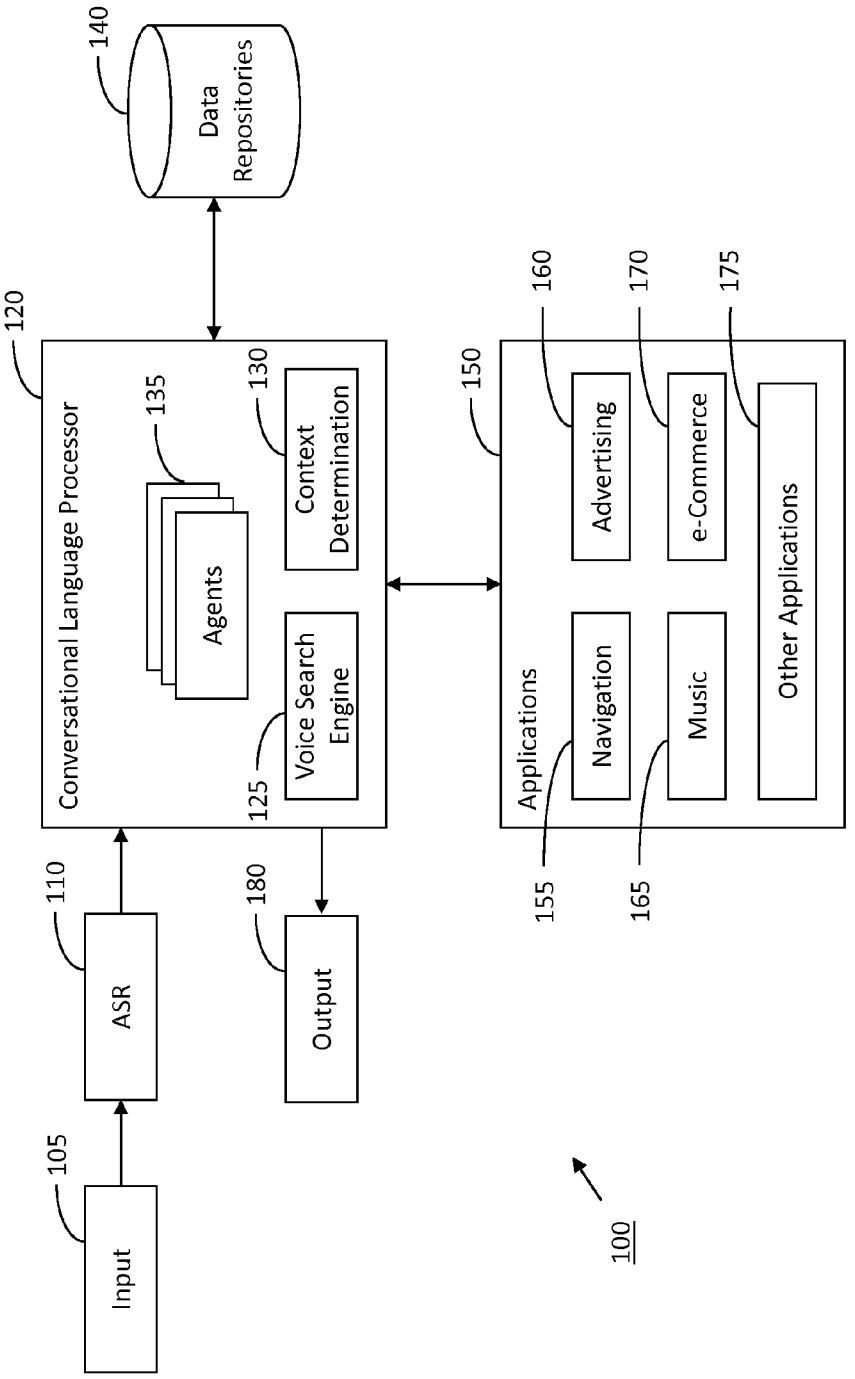

140
Data Repositories
120
Conversational Language Processor
135
Agents
130
Context Determination
125
Voice Search Engine
150
Applications
160
Advertising
170
e-Commerce
175
Other Applications
155
Navigation
165
Music
110
ASR
180
Output
105
Input
100

Figure 1

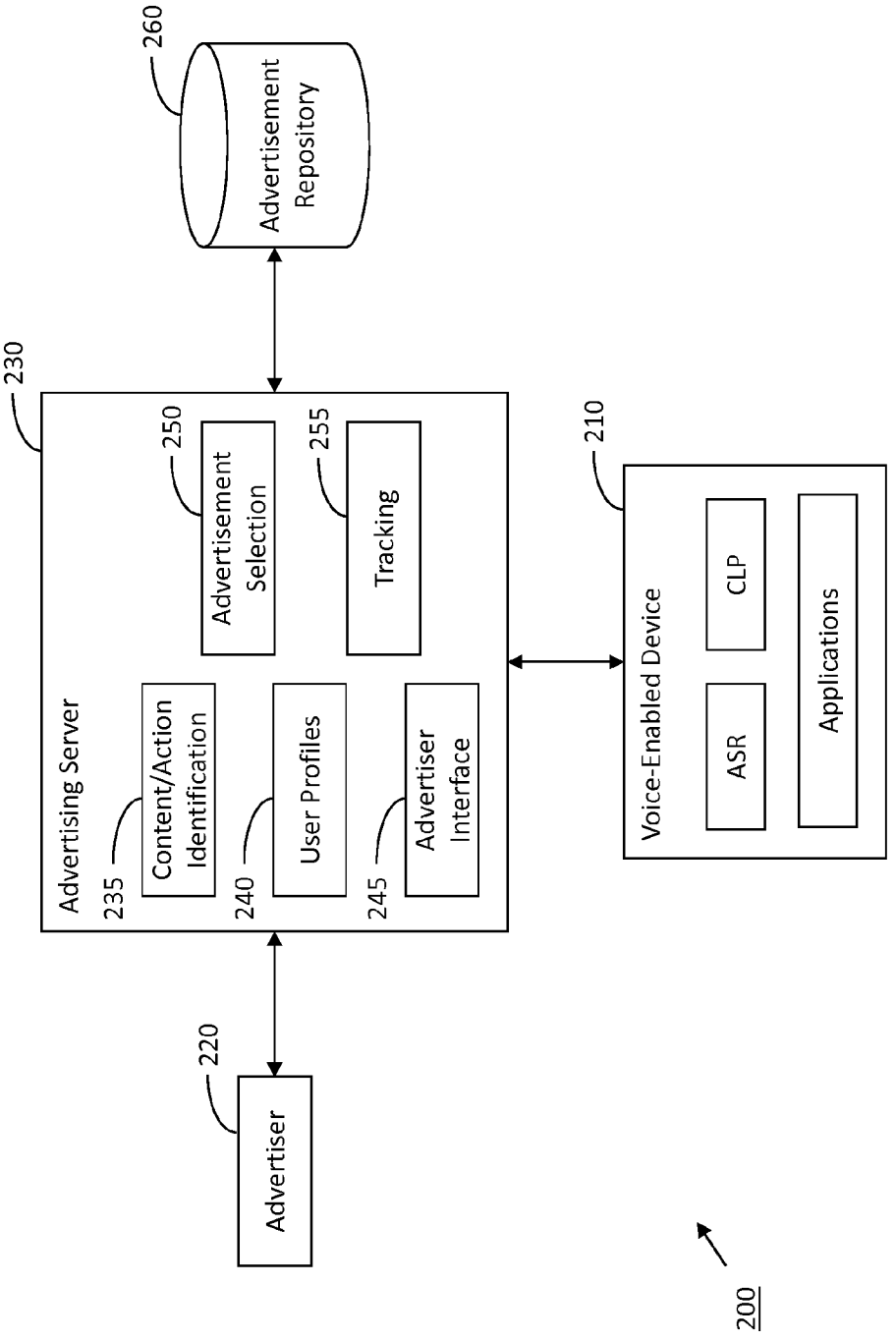

260
Advertisement Repository
230
250
Advertisement Selection
255
Tracking
210
Voice-Enabled Device
CLP
Applications
ASR
Advertising Server
Content/Action Identification
User Profiles
Advertiser Interface
235
240
245
220
Advertiser
200

Figure 2

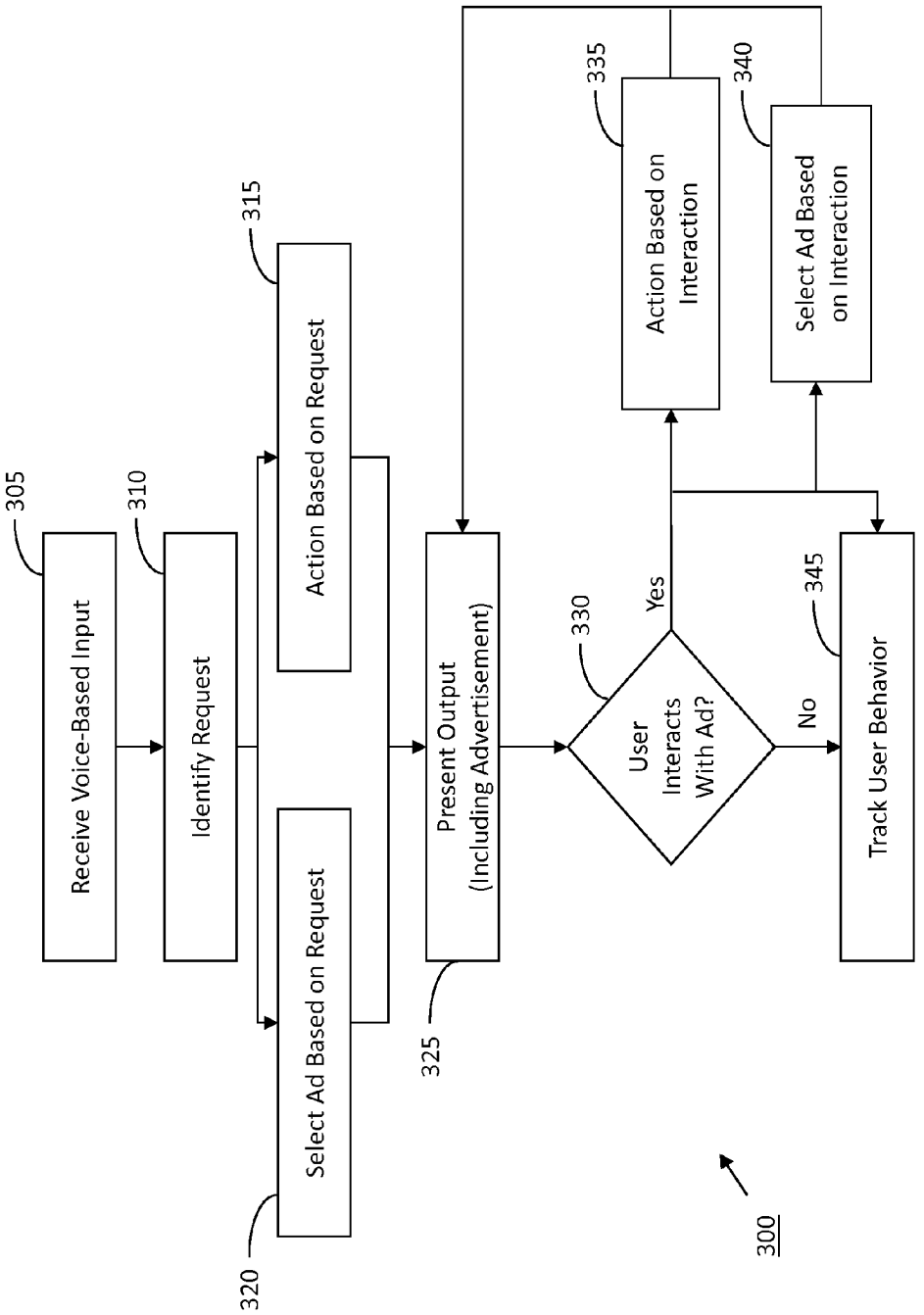

Receive Voice-Based Input
305
Identify Request
310
Action Based on Request
315
Select Ad Based on Request
320
Present Output (Including Advertisement)
325
User Interacts With Ad?
330
Yes
No
Action Based on Interaction
335
Select Ad Based on Interaction
340
Track User Behavior
345
300

Figure 3

SYSTEM AND METHOD FOR DELIVERING TARGETED ADVERTISEMENTS AND TRACKING ADVERTISEMENT INTERACTIONS IN VOICE RECOGNITION CONTEXTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of 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 selecting and presenting advertisements based on natural language processing of voice-based input.

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

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 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 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 downloads, coupons, or any other suitable marketing, advertising, campaign, or other information, as would be apparent to those skilled in the art. A voice-enabled device **210** may receive a voice-based input and establish communications with advertising server **230**. Subsequently, advertising server **230** may select one or more advertisements from among the advertisements stored in data repository **260**, and the selected advertisements may be provided to the voice-enabled device for presentation to a user.

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, geotargeting 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, 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 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 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 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 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 res-

taurant (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.

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 computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;

performing, by the one or more physical processors, speech recognition to recognize one or more words of the second natural language utterance;

determining, by the one or more physical processors, domain information for the one or more recognized words based on the first natural language utterance;

processing, by the one or more physical processors, based on the domain information, the one or more recognized words to determine an interpretation of the second natural language utterance, wherein processing the one or more recognized words comprises:

providing the one or more recognized words to a first domain agent associated with a first domain and a second domain agent associated with a second domain;

obtaining a first interpretation of the second natural language utterance from the first domain agent;

obtaining a second interpretation of the second natural language utterance from the second domain agent; and

determining the interpretation based on one or more of the first interpretation or the second interpretation;

determining, by the one or more physical processors, promotional content based on the interpretation; and

presenting, by the one or more physical processors, the promotional content to a user.

2. The method of claim **1**, wherein the first natural language utterance and the second natural language utterance are received during an interactive session between the user and the computer system, the method further comprising:

determining, by the one or more physical processors, an activity relating to one or more of the first natural language utterance or the second natural language utterance,

wherein determining the promotional content comprises determining the promotional content based on the activity.

3. The method of claim **1**, further comprising:

determining, by the one or more physical processors, context information for the one or more recognized words based on the first natural language utterance,

wherein the interpretation is determined further based on the context information.

4. The method of claim **1**, wherein the first interpretation is determined by the first domain agent based on the first domain, and wherein the second interpretation is determined by the second domain agent based on the second domain.

5. The method of claim **1**, further comprising:

associating, by the one or more physical processors, the first interpretation with a first interpretation score; and

associating, by the one or more physical processors, the second interpretation with a second interpretation score,

wherein determining the interpretation comprises determining the first interpretation as the interpretation based on the first interpretation score and the second interpretation score.

6. The method of claim **5**, wherein the first interpretation is determined as the interpretation based on the first interpretation score being greater than the second interpretation score.

7. The method of claim **1**, further comprising:

receiving, at the one or more physical processors, a third natural language utterance relating to one or more of the first natural language utterance or the second natural language utterance; and

determining, by the one or more physical processors, an interpretation of the third natural language utterance based on the promotional content.

8. The method of claim **7**, wherein the first natural language utterance, the second natural language utterance, and the third natural language utterance are received during an interactive session between the user and the computer system.

9. The method of claim **8**, wherein the third natural language utterance is received after the second natural language utterance.

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

performing, by the one or more physical processors, speech recognition to recognize one or more words of the third natural language utterance, wherein the one or more recognized words of the third natural language utterance include one or more reference words;

determining, by the one or more physical processors, that the one or more reference words refer to an item associated with the promotional content; and

initiating, by the one or more physical processors, a purchase transaction associated with the item based on the determination that the one or more reference words refer to the item.

11. The method of claim **10**, wherein the item includes a purchasable good.

12. The method of claim **7**, further comprising:

initiating, by the one or more physical processors, based on the interpretation of the third natural language utterance, a purchase transaction associated with the promotional content.

13. The method of claim **1**, wherein the first natural language utterance and the second natural language utterance are received during an interactive session between the user and the computer system, the method further comprising:

receiving, at the one or more physical processors, a third natural language utterance during the interactive session; and

processing, by the one or more physical processors, based on the third natural language utterance, the one or more recognized words to determine a reinterpretation of the second natural language utterance.

14. The method of claim **13**, wherein determining the promotional content comprises determining the promotional content based on the reinterpretation.

15. The method of claim **13**, wherein determining the promotional content comprises determining the promotional content based on the interpretation, the method further comprising:

determining, by the one or more physical processors, other promotional content based on the reinterpretation; and

presenting, by the one or more physical processors, the other promotional content to the user.

16. The method of claim **1**, further comprising:

identifying, by the one or more physical processors, one or more requests associated with the second natural language utterance,

wherein determining the promotional content comprises obtaining the promotional content based on a determination that the promotional content relates to the one or more requests.

17. The method of claim **16**, further comprising:

determining, by the one or more physical processors, one or more activities in the one or more requests; and

performing, by the one or more physical processors, the one or more activities,

wherein determining the promotional content comprises determining the promotional content based on the one or more activities.

18. The method of claim **17**, wherein determining the one or more activities comprises determining a call to be initiated for the user in response to the one or more requests, wherein performing the one or more activities comprises initiating the call for the user, and wherein determining the promotional content comprises determining the promotional content based on the call.

19. The method of claim **17**, wherein determining the one or more activities comprises determining a navigation route to be calculated for the user in response to the one or more requests, wherein performing the one or more activities comprises calculating the navigation route for the user, and wherein determining the promotional content comprises determining the promotional content based on the navigation route.

20. The method of claim **17**, wherein determining the one or more activities comprise determining a media item to be presented in response to the one or more requests, wherein performing the one or more activities comprises presenting the media item to the user, and wherein determining the promotional content comprises determining the promotional content based on the media item, the promotional content being different than the media item.

21. The method of claim 17, wherein determining the one or more activities comprises determining a reservation to be made on behalf of the user in response to the one or more requests, and wherein determining the promotional content comprises determining the promotional content based on information relating to the reservation.

22. The method of claim 17, wherein determining the one or more activities comprises determining a first reservation to be made on behalf of the user in response to the one or more requests, and wherein determining the promotional content comprises determining the promotional content based on information relating to a second reservation that is an alternative to the first reservation.

23. The method of claim 17, further comprising:

receiving, at the one or more physical processors, a third natural language utterance relating to one or more of the first natural language utterance or the second natural language utterance;

determining, by the one or more physical processors, an interpretation of the third natural language utterance based on the promotional content; and

initiating, by the one or more physical processors, based on the interpretation of the third natural language utterance, a purchase transaction associated with the promotional content.

24. The method of claim 16, further comprising:

determining, by the one or more physical processors, that at least one request of the one or more requests is incomplete or ambiguous;

monitoring, by the one or more physical processors, interaction of the user with the promotional content; and

interpreting, by the one or more physical processors, the at least one incomplete or ambiguous request based on the interaction.

25. The method of claim 16, wherein the one or more requests include a first request associated with a first application and a second request associated with a second application different than the first application, and wherein the promotional content relates to one or more of the first request or the second request.

26. The method of claim 25, wherein a first application type of the first application includes one or more of a navigation application, a music application, a commerce application, or a calling application, and wherein the second application is of an application type different than the first application type.

27. The method of claim 16, wherein the one or more requests include a first request to be processed by a first device associated with the user and a second request to be processed by a second device associated with the user, and wherein the promotional content relates to one or more of the first request or the second request.

28. The method of claim 27, wherein a first device type of the first device includes one or more of a mobile phone, a navigation device, or a media player device, and wherein the second device is of a device type different than the first device type.

29. The method of claim 27, wherein the first device and the second device are operably independent of one another.

30. The method of claim 1, further comprising:

obtaining, by the one or more physical processors, user profile information associated with the user, wherein the user profile information specifies prior user interactions with items; and

identifying, by the one or more physical processors, categories of items based on the prior user interactions specified by the user profile information,

wherein determining the promotional content comprises determining a promotional item associated with one of the categories.

31. The method of claim 1, further comprising:

obtaining, by the one or more physical processors, user profile information associated with the user, wherein the user profile information specifies prior user interactions with items;

identifying, by the one or more physical processors, one or more requests associated with the first natural language utterance or the second natural language utterance;

determining, by the one or more physical processors, one or more applications for processing the one or more requests; and

identifying, by the one or more physical processors, categories of items based on the prior user interactions specified by the user profile information, wherein the categories relate to the one or more applications,

wherein determining the promotional content comprises determining a promotional item associated with one of the categories.

32. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance;

perform speech recognition to recognize one or more words of the second natural language utterance;

determine domain information for the one or more recognized words based on the first natural language utterance;

process, based on the domain information, the one or more recognized words to determine an interpretation of the second natural language utterance, wherein processing the one or more recognized words comprises:

providing the one or more recognized words to a first domain agent associated with a first domain and a second domain agent associated with a second domain;

obtaining a first interpretation of the second natural language utterance from the first domain agent;

obtaining a second interpretation of the second natural language utterance from the second domain agent; and

determining the interpretation based on one or more of the first interpretation or the second interpretation;

determine promotional content based on the interpretation; and

present the promotional content to a user.

33. A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance during an interactive session between a user and the computer system;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance during the interactive session;

performing, by the one or more physical processors, speech recognition to recognize one or more words of the second natural language utterance;

processing, by the one or more physical processors, based on the first natural language utterance, the one or more recognized words to determine an interpretation of the second natural language utterance;

receiving, at the one or more physical processors, a third natural language utterance during the interactive session;

processing, by the one or more physical processors, based on the third natural language utterance, the one or more recognized words to determine a reinterpretation of the second natural language utterance;

determining, by the one or more physical processors, promotional content based on one or more of the interpretation or the reinterpretation; and

presenting, by the one or more physical processors, the promotional content to the user.

34. The method of claim **33**, wherein determining the promotional content comprises determining the promotional content based on the reinterpretation.

35. The method of claim **33**, wherein determining the promotional content comprises determining the promotional content based on the interpretation, the method further comprising:

determining, by the one or more physical processors, other promotional content based on the reinterpretation; and

presenting, by the one or more physical processors, the other promotional content to the user.

36. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance during an interactive session between a user and the system;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance during the interactive session;

perform speech recognition to recognize one or more words of the second natural language utterance;

process, based on the first natural language utterance, the one or more recognized words to determine an interpretation of the second natural language utterance;

receive a third natural language utterance during the interactive session;

process, based on the third natural language utterance, the one or more recognized words to determine a reinterpretation of the second natural language utterance;

determine promotional content based on one or more of the interpretation or the reinterpretation; and

present the promotional content to the user.

37. A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;

identifying, by the one or more physical processors, one or more requests associated with the second natural language utterance;

determining, by the one or more physical processors, that at least one request of the one or more requests is incomplete or ambiguous;

determining, by the one or more physical processors, promotional content that relates to the one or more requests;

presenting, by the one or more physical processors, the promotional content to a user;

monitoring, by the one or more physical processors, interaction of the user with the promotional content; and

interpreting, by the one or more physical processors, the at least one incomplete or ambiguous request based on the interaction.

38. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance;

identify one or more requests associated with the second natural language utterance;

determine that at least one request of the one or more requests is incomplete or ambiguous;

determine promotional content that relates to the one or more requests;

present the promotional content to a user;

monitor interaction of the user with the promotional content; and

interpret the at least one incomplete or ambiguous request based on the interaction.

39. A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;

identifying, by the one or more physical processors, requests associated with the second natural language utterance, wherein the requests include a first request associated with a first application and a second request associated with a second application different than the first application;

determining, by the one or more physical processors, promotional content that relates to one or more of the first request or the second request; and

presenting, by the one or more physical processors, the promotional content to a user.

40. The method of **39**, wherein a first application type of the first application includes one or more of a navigation application, a music application, a commerce application, or a calling application, and wherein the second application is of an application type different than the first application type.

41. The method of claim **39**, further comprising:

receiving, at the one or more physical processors, a third natural language utterance relating to one or more of the first natural language utterance or the second natural language utterance; and

determining, by the one or more physical processors, an interpretation of the third natural language utterance based on the promotional content.

42. The method of claim **41**, further comprising:

initiating, by the one or more physical processors, based on the interpretation of the third natural language utterance, a purchase transaction associated with the promotional content.

43. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance;

identify requests associated with the second natural language utterance, wherein the requests include a first request associated with a first application and a second request associated with a second application different than the first application;

determine promotional content that relates to one or more of the first request or the second request; and

present the promotional content to a user.

44. A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;

identifying, by the one or more physical processors, requests associated with the second natural language utterance, wherein the requests include a first request to be processed by a first device associated with a user and a second request to be processed by a second device associated with the user;

determining, by the one or more physical processors, promotional content that relates to one or more of the first request or the second request; and

presenting, by the one or more physical processors, the promotional content to the user.

45. The method of claim **44**, wherein a first device type of the first device includes one or more of a mobile phone, a navigation device, or a media player device, and wherein the second device is of a device type different than the first device type.

46. The method of claim **44**, wherein the first device and the second device are operably independent of one another.

47. The method of claim **44**, further comprising:

receiving, at the one or more physical processors, a third natural language utterance relating to one or more of the first natural language utterance or the second natural language utterance; and

determining, by the one or more physical processors, an interpretation of the third natural language utterance based on the promotional content.

48. The method of claim **47**, further comprising:

initiating, by the one or more physical processors, based on the interpretation of the third natural language utterance, a purchase transaction associated with the promotional content.

49. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance;

identify requests associated with the second natural language utterance, wherein the requests include a first request to be processed by a first device associated with a user and a second request to be processed by a second device associated with the user;

determine promotional content that relates to one or more of the first request or the second request; and

present the promotional content to the user.

50. A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;

obtaining, by the one or more physical processors, user profile information associated with a user, wherein the user profile information specifies prior user interactions with items;

identifying, by the one or more physical processors, categories of items based on the prior user interactions specified by the user profile information;

determining, by the one or more physical processors, based on the first natural language utterance and the second natural language utterance, promotional content associated with one of the categories; and

presenting, by the one or more physical processors, the promotional content to the user.

51. The method of claim **50**, further comprising:

receiving, at the one or more physical processors, a third natural language utterance relating to one or more of the first natural language utterance or the second natural language utterance; and

determining, by the one or more physical processors, an interpretation of the third natural language utterance based on the promotional content.

52. The method of claim **51**, further comprising:

initiating, by the one or more physical processors, based on the interpretation of the third natural language utterance, a purchase transaction associated with the promotional content.

53. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance;

obtain user profile information associated with a user, wherein the user profile information specifies prior user interactions with items;

identify categories of items based on the prior user interactions specified by the user profile information;

determine, based on the first natural language utterance and the second natural language utterance, promotional content associated with one of the categories; and

present the promotional content to the user.

54. A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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, at the one or more physical processors, a first natural language utterance;

providing, by the one or more physical processors, a response to the first natural language utterance;

receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;

obtaining, by the one or more physical processors, user profile information associated with a user, wherein the user profile information specifies prior user interactions with items;

identifying, by the one or more physical processors, one or more requests associated with one or more of the first natural language utterance or the second natural language utterance;

determining, by the one or more physical processors, one or more applications for processing the one or more requests; and

identifying, by the one or more physical processors, categories of items based on the prior user interactions specified by the user profile information, wherein the categories relate to the one or more applications;

determining, by the one or more physical processors, based on the first natural language utterance and the second natural language utterance, promotional content associated with one of the categories; and

presenting, by the one or more physical processors, the promotional content to the user.

55. A system for providing promotional content related to one or more natural language utterances and/or responses, the system comprising:

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

receive a first natural language utterance;

provide a response to the first natural language utterance;

receive a second natural language utterance relating to the first natural language utterance;

obtain user profile information associated with a user, wherein the user profile information specifies prior user interactions with items;

identify one or more requests associated with one or more of the first natural language utterance or the second natural language utterance;

determine one or more applications for processing the one or more requests; and

identify categories of items based on the prior user interactions specified by the user profile information, wherein the categories relate to the one or more applications;

determine, based on the first natural language utterance and the second natural language utterance, promotional content associated with one of the categories; and

present the promotional content to the user.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 8,886,536 B2
APPLICATION NO. : 14/016757
DATED : November 11, 2014
INVENTOR(S) : Tom Freeman et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the title page, in Item (72) Inventors,

change "Mike Kennwick" to --Mike Kennewick--.

Signed and Sealed this
Twenty-eighth Day of April, 2015

Michelle K. Lee

Michelle K. Lee
Director of the United States Patent and Trademark Office

EXHIBIT L

U.S. Pat. No. 8,886,536 ("the '536 patent")
Infringement Chart for Claim 44

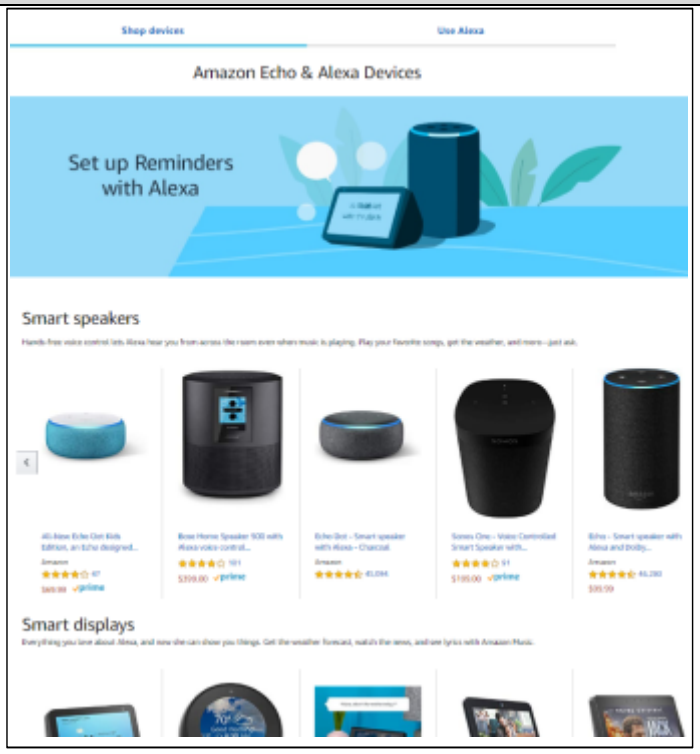
The following chart provides exemplary infringement allegations with respect to an exemplary claim of the '536 patent based on Plaintiff's investigation to date. Plaintiff reserves the right to provide modified and/or more detailed allegations (including with additional claims, products, and evidence) in accordance with its discovery obligations in view of any applicable scheduling order. This chart uses "Alexa Products" to refer to the full scope of instrumentalities identified in the complaint. Any discussion of specific Alexa Products in the following chart is exemplary.

Exemplary information relevant to infringement:

1. *Amazon Echo & Alexa Devices* - Amazon Official Site, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/?ie=UTF8&node=9818047011&ref_=sv_devicesubnav_1 (last visited Jul. 22, 2019)
2. *Alexa and Alexa Device FAQs*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (last visited Jul. 22, 2019)
3. *Place Orders with Alexa*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201807210 (last visited Jul. 22, 2019)
4. Chaim Gartenberg, *How to shop on Amazon using Alexa voice controls*, The Verge, Jul. 10, 2017, https://www.theverge.com/2017/7/10/15947672/amazon-alexa-voice-controls-shopping-prime-echo-how-to ("*How to Shop*")
5. Taylor Martin, *How to shop with Alexa on Prime Day*, CNET, Jul. 15, 2018, https://www.cnet.com/how-to/how-to-shop-with-alexa-prime-day/ ("*Prime Day*")
6. *Designing for Conversation – Build for dialog*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/ask-resources/guided/conversational-design-workshop#/lessons/gDcQVF9nOFOtNljrELzARIGN4HuD7LNh?_k=q4qp1a (last visited Jul. 22, 2019)
7. *Amazon Echo Teardown*, IFIXIT, Dec. 16, 2014, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953
8. *What is Conversational AI?*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/alexa-skills-kit/conversational-ai (last visited Jul. 22, 2019)
9. *Speech Recognizer Interface*, Alexa Voice Service, developer.amazon.com, https://developer.amazon.com/docs/alexa-voice-service/speechrecognizer.html (last visited Jul. 22, 2019)

'536 Patent Claim Language	**Exemplary Alexa Products**
44.[pr] A computer-implemented method of providing promotional content related to one or more natural language utterances and/or responses, 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:	Alexa Products are programmed to perform and have performed methods of providing promotional content related to one or more natural language utterances and/or responses, 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 recited method. For example, Alexa Products (e.g. Echo, Echo Dot, Alexa apps, Shopping apps, and Alexa cloud) include Alexa.

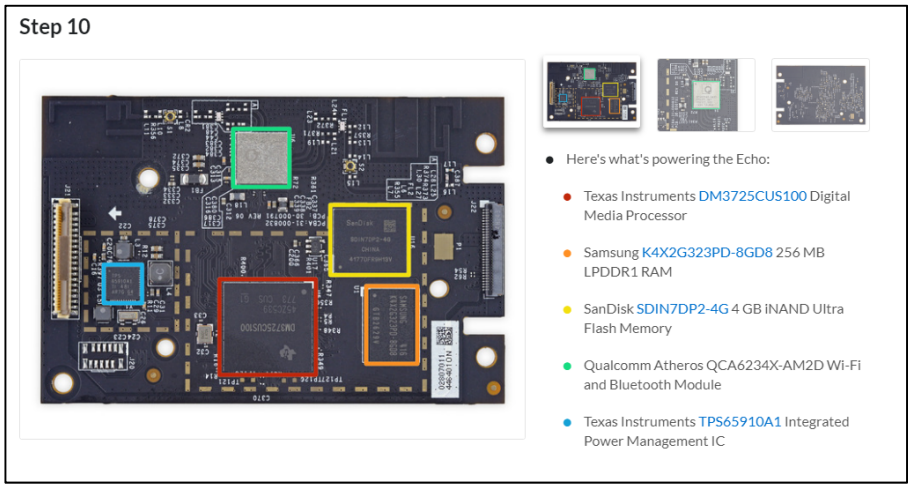
'536 Patent Claim Language	Exemplary Alexa Products



E.g. *Amazon Echo & Alexa Devices*.

"With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?*

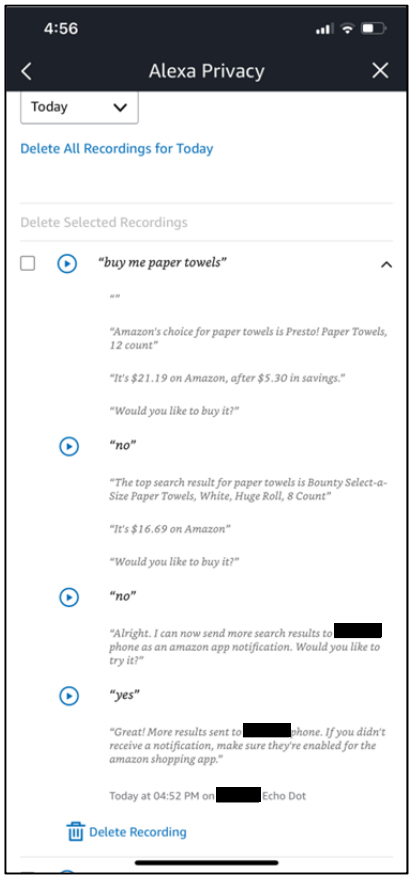
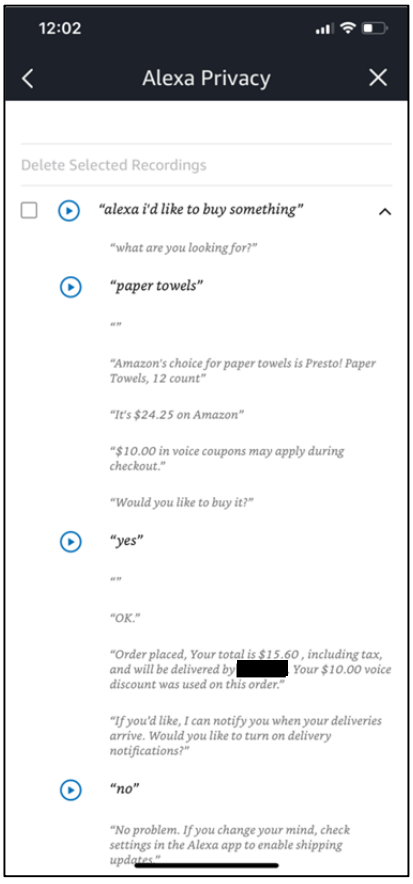
For example, the Echo has the physical processors shown below:

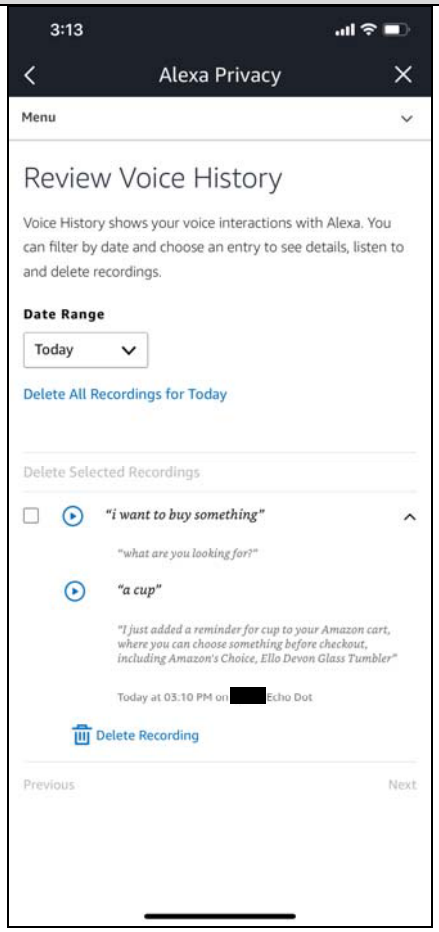


E.g. *Amazon Echo Teardown* at Step 10.

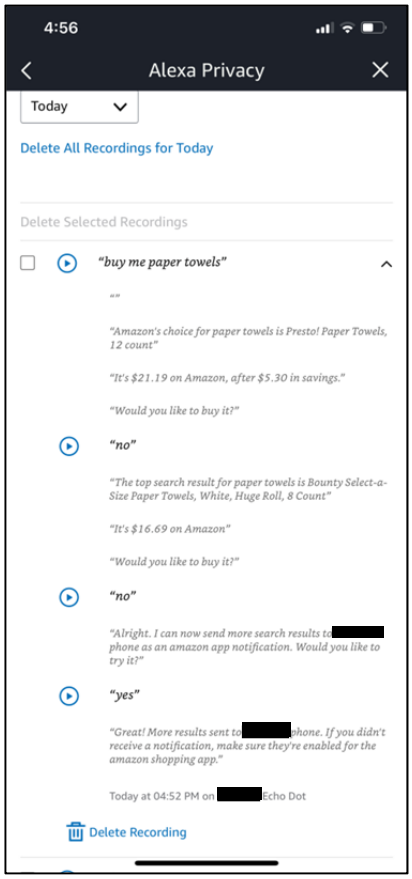
Below are more examples.

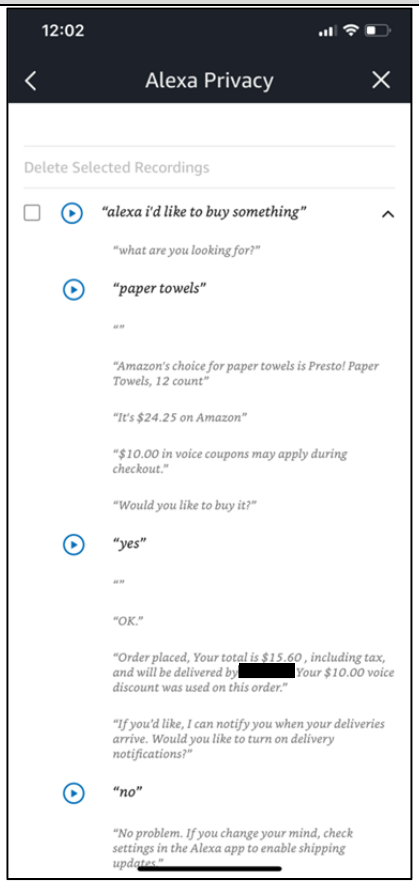
Alexa Products include Amazon's Alexa cloud. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer

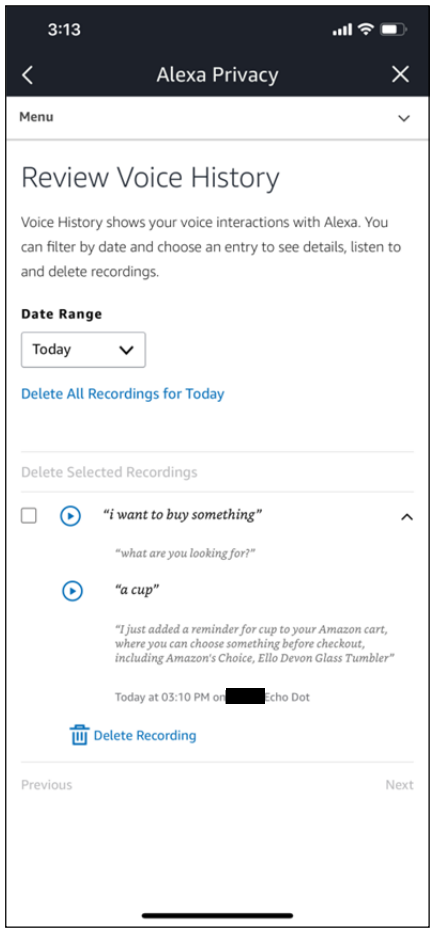
'536 Patent Claim Language	Exemplary Alexa Products
	to question 1. When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*. "You can ask Alexa to place orders for Prime-eligible physical products. On supported devices, you can also ask Alexa to place orders for music." *Place Orders with Alexa*.
[44.a] receiving, at the one or more physical processors, a first natural language utterance;	Alexa Products receive, at the one or more physical processors, a first natural language utterance. Examples of receiving a first utterance are shown in screenshots of Alexa app's "Review Voice History" page below.  

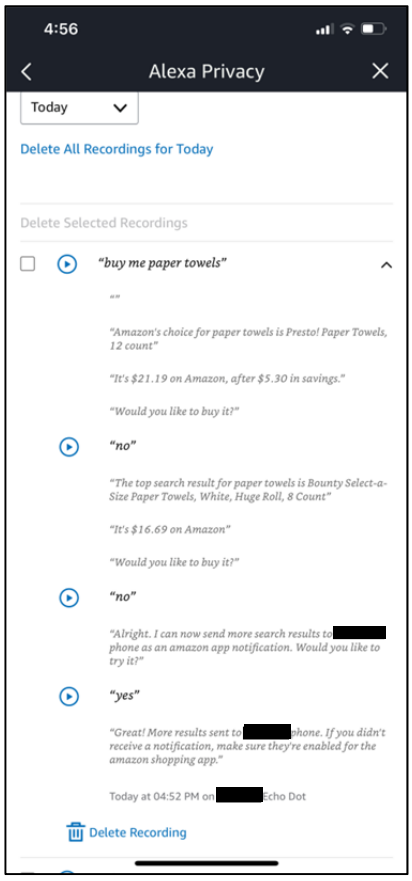
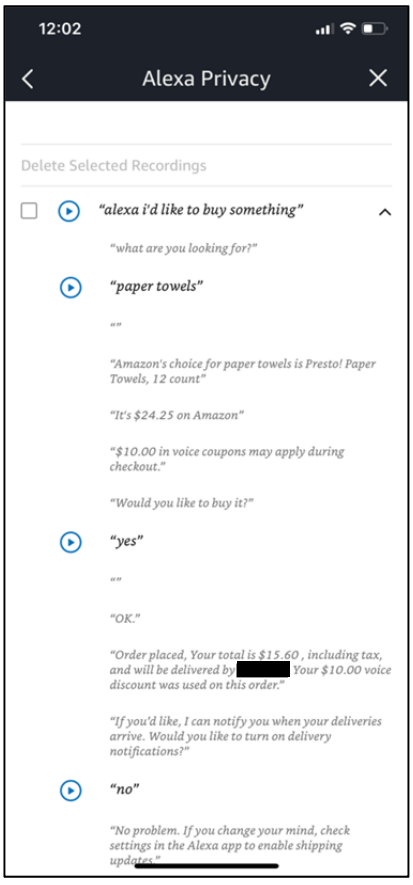
'536 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1. "Each line of the script spoken by the user is known as an utterance. The utterances are then mapped to an intent. When the user speaks one of them the associated intent will be invoked." *Designing for Conversation – Build for dialog* at Step 2. "Every user utterance leverages SpeechRecognizer. It is the core interface of the Alexa Voice Service (AVS). It exposes directives and events for capturing user speech and prompting a client when Alexa needs additional speech input." *Speech Recognizer Interface*.
[44.b] providing, by the one or more physical processors, a response to the first natural language utterance;	Alexa Products provide, by the one or more physical processors, a response to the first natural language utterance. Examples of providing a response are shown in screenshots of Alexa app's "Review Voice History" page below.

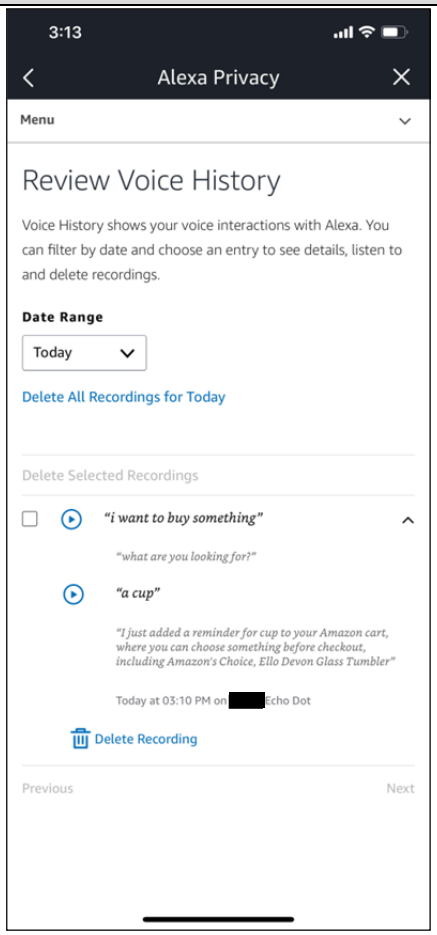
’536 Patent Claim Language	Exemplary Alexa Products

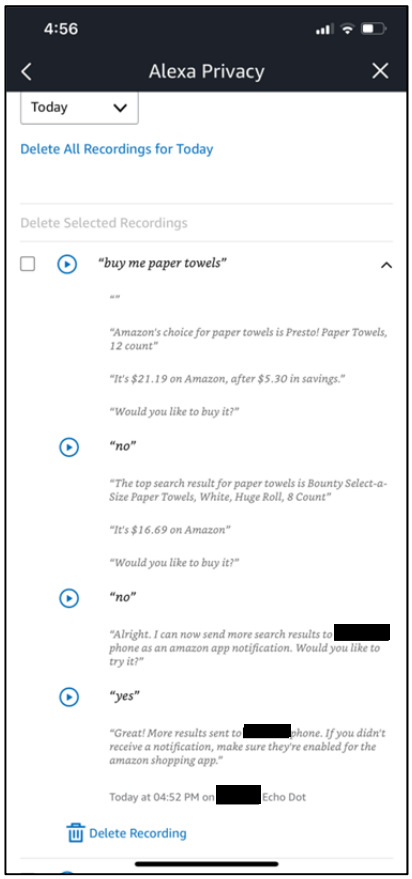
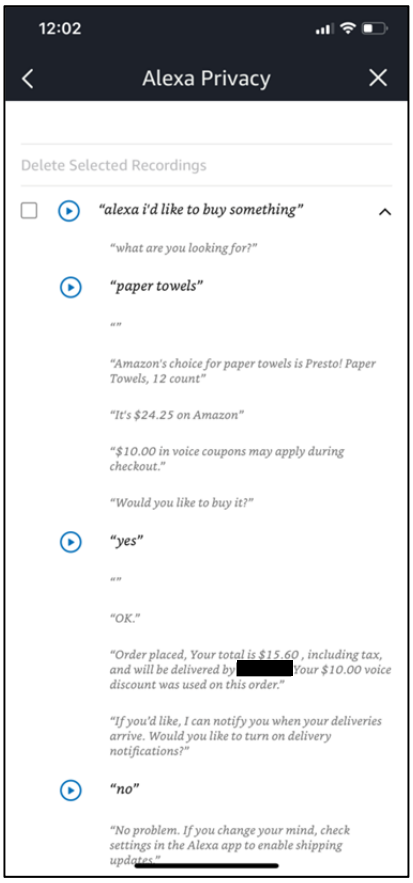




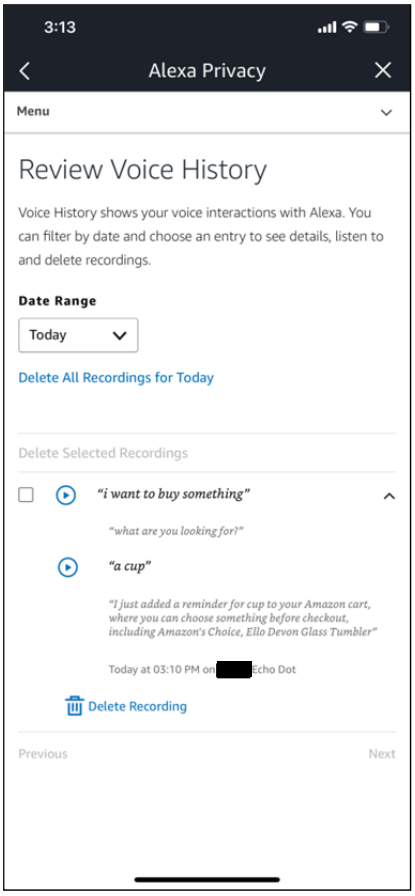


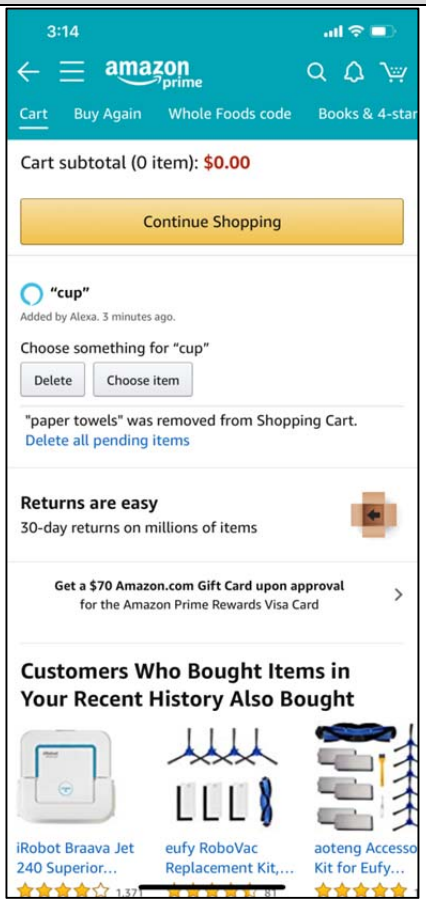
’536 Patent Claim Language	Exemplary Alexa Products
	Below are more examples. “It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add in your cart.” *Prime Day*. “Conversations are not isolated occurrences. They build on previous questions and answers, whether from five minutes or five days ago. When appropriate, voice-first experiences should account for prior context to make the conversation more engaging or relevant to the user.” *What is Conversational AI?* “Every user utterance leverages SpeechRecognizer. It is the core interface of the Alexa Voice Service (AVS). It exposes directives and events for capturing user speech and prompting a client when Alexa needs additional speech input.” *Speech Recognizer Interface*.
[44.c] receiving, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance;	Alexa Products, receive, at the one or more physical processors, a second natural language utterance relating to the first natural language utterance. Examples of receiving a second utterance are shown in screenshots of Alexa app’s “Review Voice History” page below.  

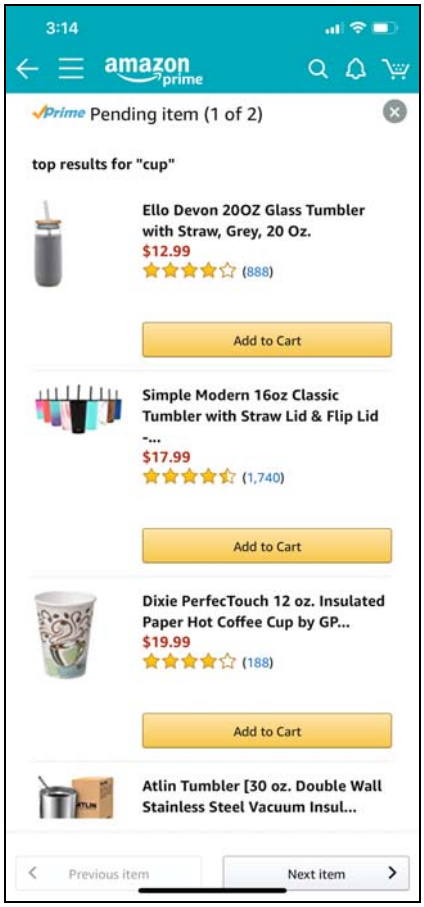
'536 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add in your cart." *Prime Day*. "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1. "Each line of the script spoken by the user is known as an utterance. The utterances are then mapped to an intent. When the user speaks one of them the associated intent will be invoked." *Designing for Conversation – Build for dialog* at Step 2. "Every user utterance leverages SpeechRecognizer. It is the core interface of the Alexa Voice Service (AVS). It exposes directives and events for capturing user speech and prompting a client when Alexa needs additional speech input." *Speech Recognizer Interface*.
[44.d] identifying, by the one or more physical processors, requests associated with the second natural	Alexa Products identify, by the one or more physical processors, requests associated with the second natural language utterance, wherein the requests include a first request to be processed by a first device

'536 Patent Claim Language	Exemplary Alexa Products
language utterance, wherein the requests include a first request to be processed by a first device associated with a user and a second request to be processed by a second device associated with the user;	associated with a user and a second request to be processed by a second device associated with the user. Examples are identifying requests associated with the second natural language utterance, as shown below in screenshots of Alexa app's "Review Voice History" and Amazon app's "cart" and "results" pages below.  

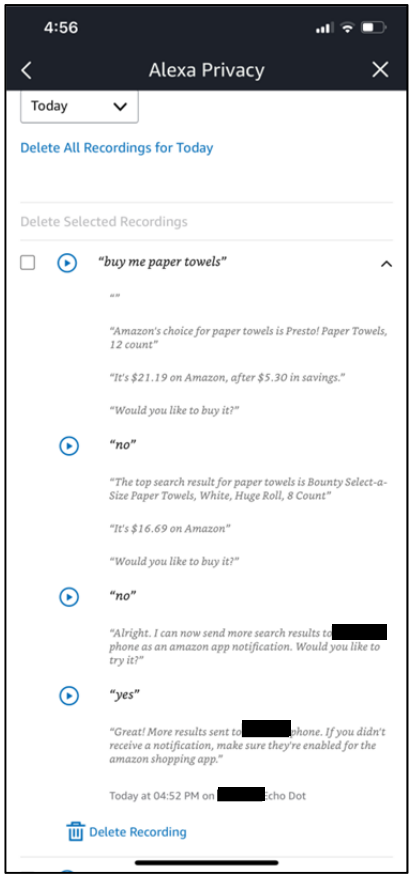
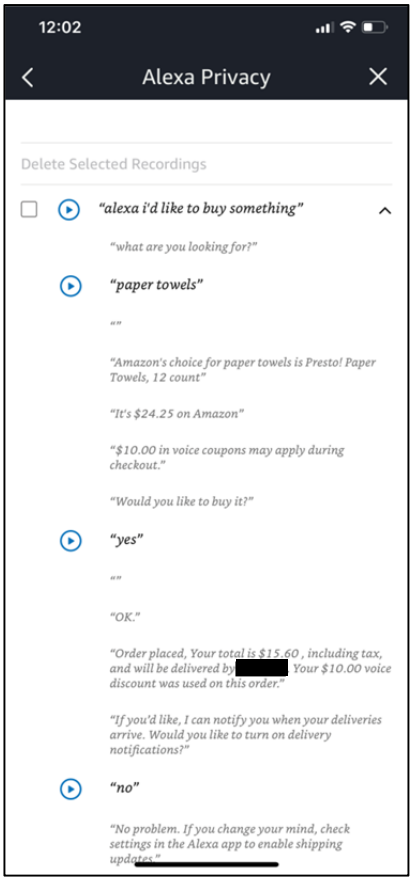
'536 Patent Claim Language	Exemplary Alexa Products



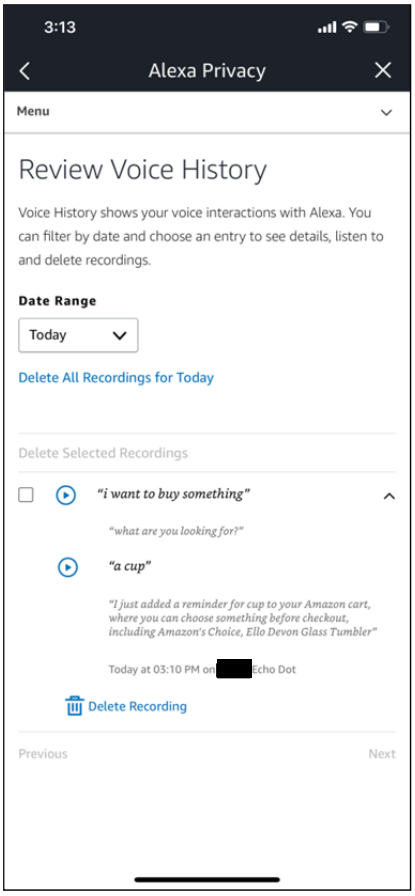


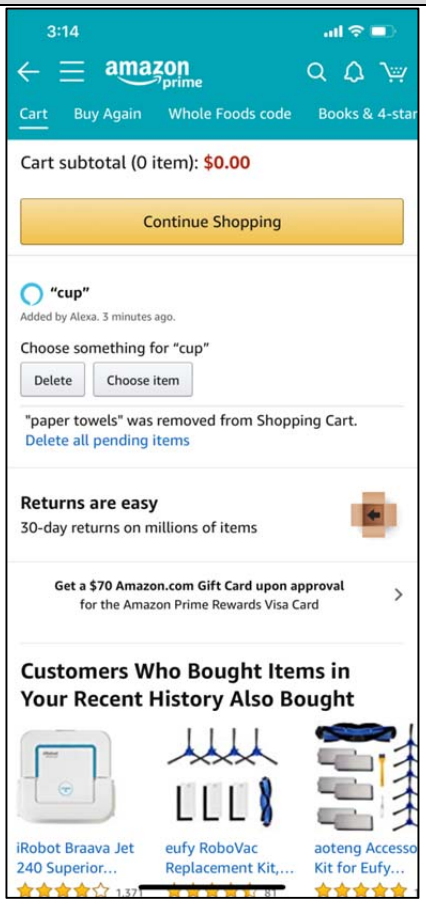


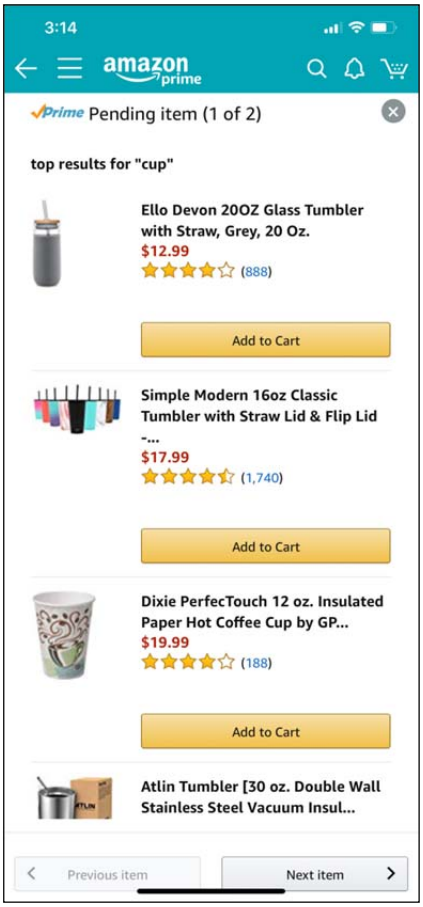
Another example is: "When you speak to Alexa, a recording of what

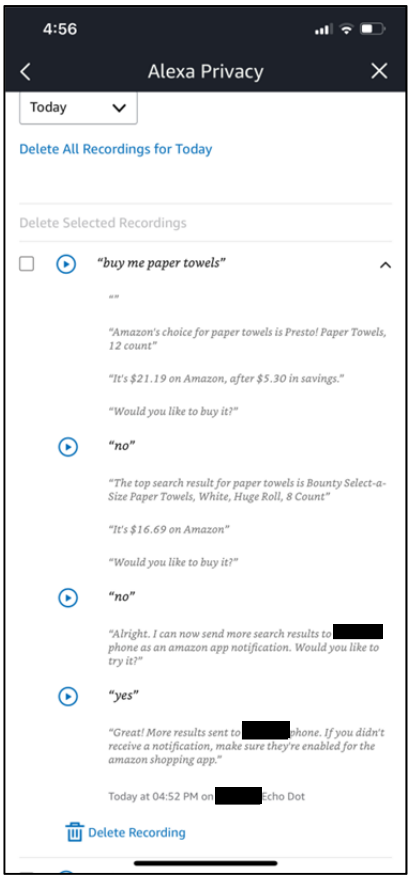
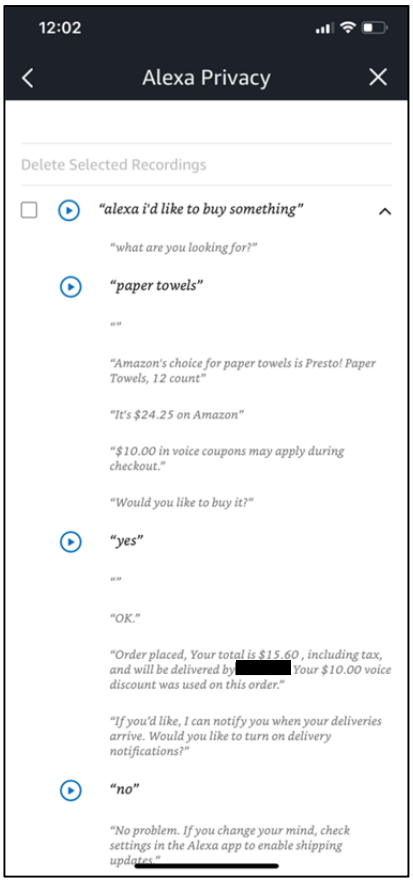
'536 Patent Claim Language	Exemplary Alexa Products
	you asked Alexa is sent to Amazon's cloud where we process your request and other information to respond to you." *Alexa and Alexa Device FAQs* at answer to question 2.
[44.e] determining, by the one or more physical processors, promotional content that relates to one or more of the first request or the second request; and	Alexa Products determine, by the one or more physical processors, promotional content that relates to one or more of the first request or the second request. Examples are determining promotional content, as shown below in screenshots of Alexa app's "Review Voice History" and Amazon app's "cart" and "results" pages below.  

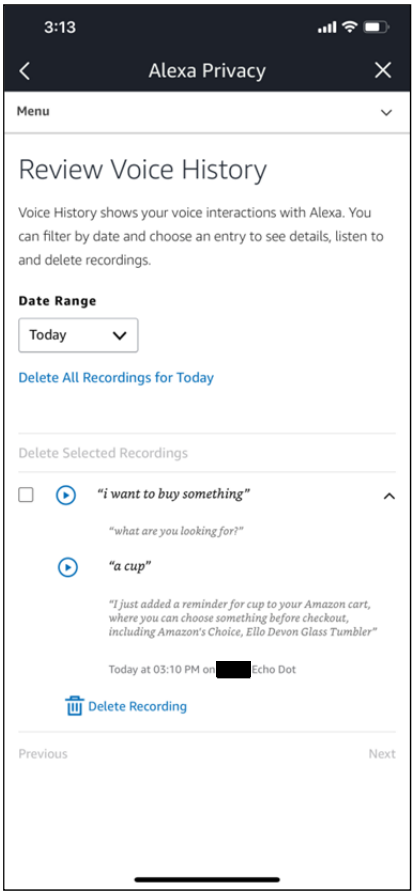
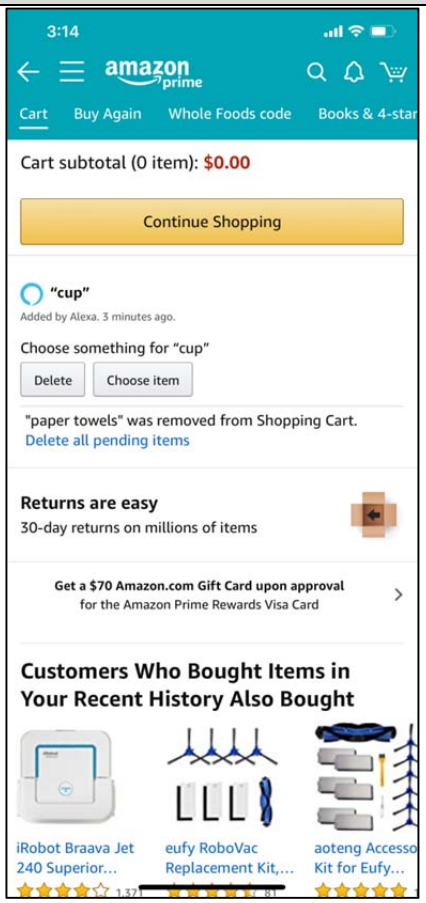
’536 Patent Claim Language	Exemplary Alexa Products







'536 Patent Claim Language	Exemplary Alexa Products
	More examples are below. When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*. "Conversations are not isolated occurrences. They build on previous questions and answers, whether from five minutes or five days ago. When appropriate, voice-first experiences should account for prior context to make the conversation more engaging or relevant to the user." *What is Conversational AI?*
[44.f] presenting, by the one or more physical processors, the promotional content to the user.	Alexa Products present, by the one or more physical processors, the promotional content to the user. Examples of presenting the promotional content are shown below in screenshots of Alexa app's "Review Voice History" and Amazon app's "cart" and "results" pages below.  

’536 Patent Claim Language	Exemplary Alexa Products
	 

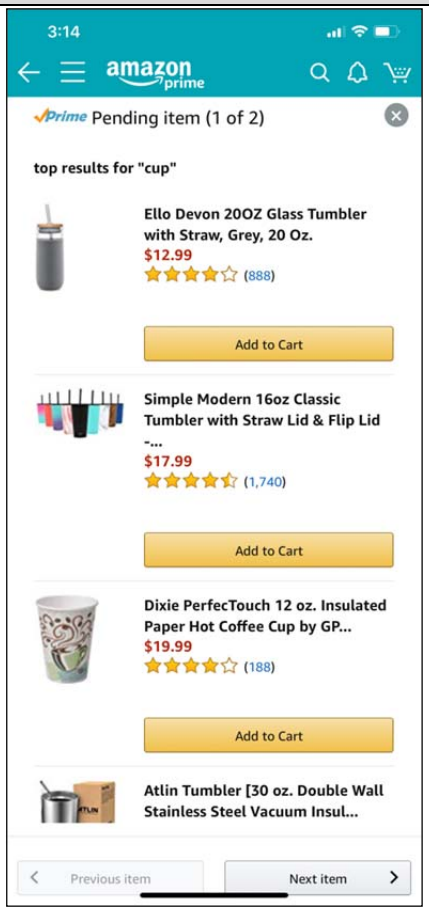
'536 Patent Claim Language	Exemplary Alexa Products
	 Below are more examples. "It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add in your cart." *Prime Day*. When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*.

EXHIBIT M

US009269097B2

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

(10) **Patent No.: US 9,269,097 B2**
(45) **Date of Patent: Feb. 23, 2016**

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

(71) Applicant: **VOICEBOX TECHNOLOGIES CORPORATION**, Bellevue, WA (US)

(72) Inventors: **Tom Freeman**, Mercer Island, WA (US); **Mike Kennewick**, Bellevue, 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 0 days.

(21) Appl. No.: **14/537,598**

(22) Filed: **Nov. 10, 2014**

(65) **Prior Publication Data**

US 2015/0066627 A1 Mar. 5, 2015

Related U.S. Application Data

(63) 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

(Continued)

(51) **Int. Cl.**
G10L 15/18 (2013.01)
G06Q 30/02 (2012.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G06Q 30/0251*** (2013.01); ***G06F 17/275*** (2013.01); ***G06Q 30/0241*** (2013.01); ***G06Q 30/0242*** (2013.01); ***G10L 15/18*** (2013.01); ***G10L 15/26*** (2013.01); ***G10L 15/265*** (2013.01)

(58) **Field of Classification Search**
CPC G10L 15/18
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

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

FOREIGN PATENT DOCUMENTS

EP 1 320 043 A2 6/2003
EP 1 646 037 4/2006
(Continued)

OTHER PUBLICATIONS

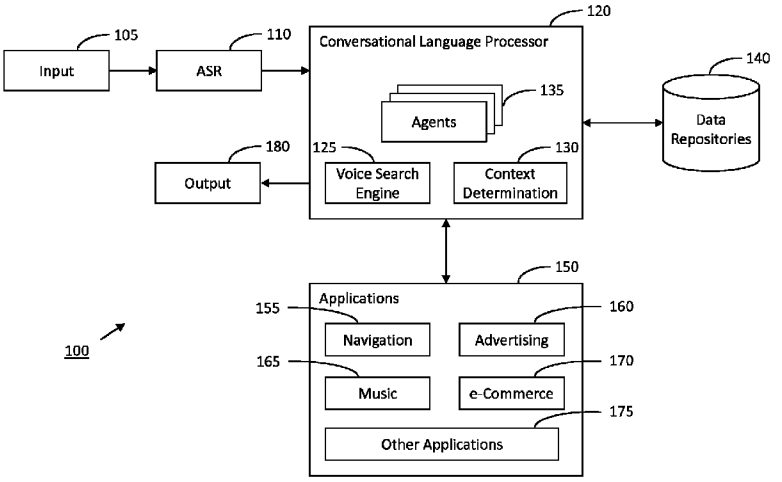
Reuters, "IBM to Enable Honda Drivers to Talk to Cars", Charles Schwab & Co., Inc., Jul. 28, 2002, 1 page.
(Continued)

Primary Examiner — Brian 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.

23 Claims, 3 Drawing Sheets



Related U.S. Application Data

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.**

G10L 15/26	(2006.01)
G06F 17/27	(2006.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,829,423	A	5/1989	Tennant et al.	364/200
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
5,274,560	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
5,475,733	A	12/1995	Eisdorfer et al.	379/52
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
5,617,407	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
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 et al.	395/2.66
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
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
5,832,221	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
5,867,817	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
5,902,347	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 et al.	707/1
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
5,960,397	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
5,987,404	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
6,076,059	A	6/2000	Glickman et al.	704/260
6,078,886	A	6/2000	Dragosh et al.	704/270
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
6,144,938	A	11/2000	Surace et al.	704/257
6,154,526	A	11/2000	Dahlke et al.	379/88.03
6,160,883	A	12/2000	Jackson et al.	379/230
6,167,377	A	12/2000	Gillick et al.	704/240
6,173,266	B1	1/2001	Marx et al.	704/270
6,173,279	B1	1/2001	Levin et al.	707/5
6,175,858	B1	1/2001	Bulfer et al.	709/206
6,185,535	B1	2/2001	Hedin et al.	704/270
6,188,982	B1	2/2001	Chiang	704/256
6,192,110	B1	2/2001	Abella et al.	379/88.01
6,192,338	B1	2/2001	Haszto et al.	704/257
6,195,634	B1	2/2001	Dudemaine et al.	704/231
6,195,651	B1	2/2001	Handel et al.	707/2
6,199,043	B1	3/2001	Happ	704/272
6,208,964	B1	3/2001	Sabourin	704/244
6,208,972	B1	3/2001	Grant et al.	704/275
6,219,346	B1	4/2001	Maxemchuk	370/338
6,219,643	B1	4/2001	Cohen et al.	704/257
6,226,612	B1	5/2001	Srenger et al.	704/256
6,233,556	B1	5/2001	Teunen et al.	704/250
6,233,559	B1	5/2001	Balakrishnan	704/275
6,233,561	B1	5/2001	Junqua et al.	704/277
6,236,968	B1	5/2001	Kanevsky et al.	704/275
6,246,981	B1	6/2001	Papineni et al.	704/235
6,246,990	B1	6/2001	Happ	704/275
6,266,636	B1	7/2001	Kosaka et al.	704/244
6,269,336	B1	7/2001	Ladd et al.	704/270
6,272,455	B1	8/2001	Hoshen et al.	704/1
6,275,231	B1	8/2001	Obradovich	345/349
6,278,377	B1	8/2001	DeLine et al.	340/815.4
6,278,968	B1	8/2001	Franz et al.	704/3
6,286,002	B1	9/2001	Axaopoulos	
6,288,319	B1	9/2001	Catona	84/609
6,292,767	B1	9/2001	Jackson et al.	704/1
6,301,560	B1	10/2001	Masters	704/251
6,308,151	B1	10/2001	Smith	704/235
6,311,159	B1	10/2001	Van Tichelen et al.	704/275
6,314,402	B1	11/2001	Monaco et al.	704/275
6,321,196	B1	11/2001	Franceschi	704/243
6,356,869	B1	3/2002	Chapados et al.	704/275

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,362,748	B1	3/2002	Huang	340/901
6,366,882	B1	4/2002	Bijl et al.	704/235
6,366,886	B1	4/2002	Dragosh et al.	704/270.1
6,374,214	B1	4/2002	Friedland et al.	704/235
6,377,913	B1	4/2002	Coffman et al.	704/8
6,381,535	B1	4/2002	Durocher et al.	701/202
6,385,596	B1	5/2002	Wiser et al.	705/51
6,385,646	B1	5/2002	Brown et al.	709/217
6,393,403	B1	5/2002	Majaniemi	704/275
6,393,428	B1	5/2002	Miller et al.	707/102
6,397,181	B1	5/2002	Li et al.	704/256
6,404,878	B1	6/2002	Jackson et al.	379/221.01
6,405,170	B1	6/2002	Phillips et al.	704/270
6,408,272	B1	6/2002	White et al.	704/270.1
6,411,810	B1	6/2002	Maxemchuk	455/453
6,411,893	B2	6/2002	Ruhl	701/207
6,415,257	B1	7/2002	Junqua et al.	704/275
6,418,210	B1	7/2002	Sayko	379/142.15
6,420,975	B1	7/2002	DeLine et al.	340/815.4
6,429,813	B2	8/2002	Feigen	342/357.13
6,430,285	B1	8/2002	Bauer et al.	379/265.01
6,430,531	B1	8/2002	Polish	704/257
6,434,523	B1	8/2002	Monaco	704/257
6,434,524	B1	8/2002	Weber	704/257
6,434,529	B1	8/2002	Walker et al.	704/275
6,442,522	B1	8/2002	Carberry et al.	704/257
6,446,114	B1	9/2002	Bulfer et al.	709/206
6,453,153	B1	9/2002	Bowker et al.	455/67.4
6,453,292	B2	9/2002	Ramaswamy et al.	704/235
6,456,711	B1	9/2002	Cheung et al.	379/265.09
6,456,974	B1	9/2002	Baker et al.	704/270.1
6,466,654	B1	10/2002	Cooper et al.	379/88.01
6,466,899	B1	10/2002	Yano et al.	704/1
6,470,315	B1	10/2002	Netsch et al.	704/256
6,487,494	B2	11/2002	Odinak et al.	701/202
6,487,495	B1	11/2002	Gale et al.	701/209
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
6,505,155	B1	1/2003	Vanbuskirk et al.	704/246
6,510,417	B1	1/2003	Woods et al.	704/275
6,513,006	B2	1/2003	Howard et al.	704/257
6,522,746	B1	2/2003	Marchok et al.	379/406.03
6,523,061	B1	2/2003	Halverson et al.	709/202
6,532,444	B1	3/2003	Weber	704/257
6,539,348	B1	3/2003	Bond et al.	704/9
6,549,629	B2	4/2003	Finn et al.	381/92
6,553,372	B1	4/2003	Brassell et al.	707/5
6,556,970	B1	4/2003	Sasaki et al.	704/257
6,556,973	B1	4/2003	Lewin	704/277
6,560,576	B1	5/2003	Cohen et al.	704/270
6,560,590	B1	5/2003	Shwe et al.	706/55
6,567,778	B1	5/2003	Chao Chang et al.	704/257
6,567,797	B1	5/2003	Schuetze et al.	707/2
6,570,555	B1	5/2003	Prevost et al.	345/156
6,570,964	B1	5/2003	Murveit et al.	379/67.1
6,571,279	B1	5/2003	Herz et al.	709/217
6,574,597	B1	6/2003	Mohri et al.	704/251
6,574,624	B1	6/2003	Johnson et al.	707/5
6,578,022	B1	6/2003	Foulger et al.	706/45
6,581,103	B1	6/2003	Dengler	709/231
6,584,439	B1	6/2003	Geilhufe et al.	704/270
6,587,858	B1	7/2003	Strazza	707/102
6,591,239	B1	7/2003	McCall et al.	704/275
6,594,257	B1	7/2003	Doshi et al.	370/352
6,594,367	B1	7/2003	Marash et al.	381/92
6,598,018	B1	7/2003	Junqua	704/251
6,601,026	B2	7/2003	Appelt et al.	704/9
6,604,075	B1	8/2003	Brown et al.	704/270.1
6,604,077	B2	8/2003	Dragosh et al.	704/270.1
6,606,598	B1	8/2003	Holthouse et al.	704/275
6,611,692	B2	8/2003	Raffel et al.	455/552
6,614,773	B1	9/2003	Maxemchuk	370/337
6,615,172	B1	9/2003	Bennett et al.	704/257
6,622,119	B1	9/2003	Ramaswamy et al.	704/9
6,629,066	B1	9/2003	Jackson et al.	704/9
6,631,346	B1	10/2003	Karaorman et al.	704/9
6,631,351	B1	10/2003	Ramachandran et al.	704/270
6,633,846	B1	10/2003	Bennett et al.	704/257
6,636,790	B1	10/2003	Lightner et al.	701/33
6,643,620	B1	11/2003	Contolini et al.	704/270
6,647,363	B2	11/2003	Claassen	704/1
6,650,747	B1	11/2003	Bala et al.	379/265.06
6,658,388	B1	12/2003	Kleindienst et al.	704/275
6,678,680	B1	1/2004	Woo	707/6
6,681,206	B1	1/2004	Gorin et al.	704/243
6,691,151	B1	2/2004	Cheyer 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
6,721,633	B2	4/2004	Funk et al.	701/1
6,721,706	B1	4/2004	Strubbe et al.	704/275
6,726,636	B2	4/2004	Der Ghazarian et al.	600/532
6,735,592	B1	5/2004	Neumann et al.	707/101
6,739,556	B1	5/2004	Langston	244/189
6,741,931	B1	5/2004	Kohut et al.	701/209
6,742,021	B1	5/2004	Halverson et al.	709/218
6,745,161	B1	6/2004	Arnold et al.	704/7
6,751,591	B1	6/2004	Gorin et al.	704/257
6,751,612	B1	6/2004	Schuetze et al.	707/4
6,754,485	B1	6/2004	Obradovich et al.	455/414.1
6,754,627	B2	6/2004	Woodward	704/235
6,757,544	B2	6/2004	Rangarajan et al.	455/456.1
6,757,718	B1	6/2004	Halverson et al.	709/218
6,795,808	B1	9/2004	Strubbe et al.	704/275
6,801,604	B2	10/2004	Maes et al.	379/88.17
6,801,893	B1	10/2004	Backfried et al.	704/257
6,813,341	B1	11/2004	Mahoney	379/88.01
6,829,603	B1	12/2004	Chai et al.	707/5
6,832,230	B1	12/2004	Zilliacus et al.	707/203
6,833,848	B1	12/2004	Wolff et al.	345/719
6,850,603	B1	2/2005	Eberle et al.	379/88.16
6,856,990	B2	2/2005	Barile et al.	707/10
6,865,481	B2	3/2005	Kawazoe et al.	701/211
6,868,380	B2	3/2005	Kroeker	704/240
6,868,385	B1	3/2005	Gerson	704/275
6,873,837	B1	3/2005	Yoshioka et al.	455/321
6,877,001	B2	4/2005	Wolf et al.	707/3
6,877,134	B1	4/2005	Fuller et al.	715/500.1
6,901,366	B1	5/2005	Kuhn et al.	704/275
6,910,003	B1	6/2005	Arnold et al.	704/4
6,912,498	B2	6/2005	Stevens et al.	704/235
6,915,126	B2	7/2005	Mazzara, Jr.	455/411
6,928,614	B1	8/2005	Everhart	715/728
6,934,756	B2	8/2005	Maes	709/227
6,937,977	B2	8/2005	Gerson	704/201
6,937,982	B2	8/2005	Kitaoka et al.	704/252
6,941,266	B1	9/2005	Gorin et al.	704/257
6,944,594	B2	9/2005	Busayapongchai et al.	704/275
6,950,821	B2	9/2005	Faybishenko et al.	707/10
6,954,755	B2	10/2005	Reisman	707/10
6,959,276	B2	10/2005	Droppo et al.	704/226
6,961,700	B2	11/2005	Mitchell et al.	704/235
6,963,759	B1	11/2005	Gerson	455/563
6,964,023	B2	11/2005	Maes et al.	715/811
6,968,311	B2	11/2005	Knockeart et al.	704/270
6,973,387	B2	12/2005	Masclet et al.	701/211
6,975,993	B1	12/2005	Keiller	704/275
6,980,092	B2	12/2005	Turnbull et al.	340/425.5
6,983,055	B2	1/2006	Luo	381/313
6,990,513	B2	1/2006	Belfiore et al.	709/203
6,996,531	B2	2/2006	Korall et al.	704/270
7,003,463	B1	2/2006	Maes et al.	704/270.1
7,016,849	B2	3/2006	Arnold et al.	704/275
7,020,609	B2	3/2006	Thrift et al.	704/270.1
7,024,364	B2	4/2006	Guerra et al.	704/270
7,027,586	B2	4/2006	Bushey et al.	379/265.09
7,027,975	B1	4/2006	Pazandak et al.	704/9
7,035,415	B2	4/2006	Belt et al.	381/92

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,036,128 B1 4/2006 Julia et al. 719/317
7,043,425 B2 5/2006 Pao 704/211
7,054,817 B2 5/2006 Shao 704/270
7,058,890 B2 6/2006 George et al. 715/728
7,062,488 B1 6/2006 Reisman 707/8
7,069,220 B2 6/2006 Coffman et al. 704/275
7,072,834 B2 7/2006 Zhou 704/244
7,076,362 B2 7/2006 Ohtsuji et al. 701/211
7,082,469 B2 7/2006 Gold et al. 709/231
7,085,708 B2 8/2006 Manson 704/9
7,092,928 B1 8/2006 Elad et al. 706/60
7,107,210 B2 9/2006 Deng et al. 704/226
7,107,218 B1 9/2006 Preston 704/270
7,110,951 B1 9/2006 Lemelson et al. 704/270
7,127,395 B1 10/2006 Gorin et al. 704/257
7,127,400 B2 10/2006 Koch 704/270.1
7,130,390 B2 10/2006 Abburi 379/88.17
7,136,875 B2 11/2006 Anderson et al. 707/104.1
7,137,126 B1 11/2006 Coffman et al. 719/328
7,143,037 B1 11/2006 Chestnut 704/251
7,143,039 B1 11/2006 Stifelman et al. 704/270
7,146,319 B2 12/2006 Hunt 704/254
7,149,696 B2 12/2006 Shimizu et al. 705/1
7,165,028 B2 1/2007 Gong 704/233
7,170,993 B2 1/2007 Anderson et al. 379/265.09
7,171,291 B2 1/2007 Obradovich 701/29
7,174,300 B2 2/2007 Bush 704/275
7,177,798 B2 2/2007 Hsu et al. 704/9
7,184,957 B2 2/2007 Brookes et al. 704/246
7,190,770 B2 3/2007 Ando et al. 379/88.01
7,197,069 B2 3/2007 Agazzi et al. 375/233
7,197,460 B1 3/2007 Gupta et al. 704/270.1
7,203,644 B2 4/2007 Anderson et al. 704/246
7,206,418 B2 4/2007 Yang et al. 381/92
7,207,011 B2 4/2007 Mulvey et al. 715/812
7,215,941 B2 5/2007 Beckmann et al. 455/404.1
7,228,276 B2 6/2007 Omote et al. 704/243
7,231,343 B1 6/2007 Treadgold et al. 704/9
7,236,923 B1 6/2007 Gupta 704/9
7,254,482 B2 8/2007 Kawasaki et al. 701/211
7,272,212 B2 9/2007 Eberle et al. 379/88.17
7,277,854 B2 10/2007 Bennett et al. 704/257
7,283,829 B2 10/2007 Christenson et al. 455/461
7,283,951 B2 10/2007 Marchisio et al. 704/9
7,289,606 B2 10/2007 Sibal et al. 379/52
7,299,186 B2 11/2007 Kuzunuki et al. 704/270.1
7,301,093 B2 11/2007 Sater et al. 84/615
7,305,381 B1 12/2007 Poppink et al.
7,321,850 B2 1/2008 Wakita 704/10
7,328,155 B2 2/2008 Endo et al. 704/251
7,337,116 B2 2/2008 Charlesworth et al. 704/254
7,340,040 B1 3/2008 Saylor et al. 379/67.1
7,366,285 B2 4/2008 Parolkar et al. 379/88.17
7,366,669 B2 4/2008 Nishitani et al. 704/256
7,376,645 B2 5/2008 Bernard 707/3
7,386,443 B1 6/2008 Parthasarathy et al. 704/201
7,398,209 B2 7/2008 Kennewick et al. 704/255
7,406,421 B2 7/2008 Odinak et al. 704/275
7,415,414 B2 8/2008 Azara et al. 704/270
7,421,393 B1 9/2008 Di Fabbrizio et al. 704/275
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
7,454,608 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
7,472,075 B2 12/2008 Odinak et al. 705/26
7,477,909 B2 1/2009 Roth 455/466
7,478,036 B2 1/2009 Shen et al. 704/9
7,487,088 B1 2/2009 Gorin et al. 704/240
7,487,110 B2 2/2009 Bennett et al. 705/26
7,493,259 B2 2/2009 Jones et al. 704/257
7,493,559 B1 2/2009 Wolff et al. 715/727
7,502,672 B1 3/2009 Kolls 701/29
7,502,738 B2 3/2009 Kennewick et al. 704/257
7,516,076 B2 4/2009 Walker et al. 704/275
7,529,675 B2 5/2009 Maes 704/270.1
7,536,297 B2 5/2009 Byrd et al. 704/10
7,536,374 B2 5/2009 Au 706/55
7,542,894 B2 6/2009 Murata 704/9
7,546,382 B2 6/2009 Healey et al. 709/246
7,548,491 B2 6/2009 Macfarlane 367/198
7,552,054 B1 6/2009 Stifelman et al. 704/270
7,558,730 B2 7/2009 Davis et al. 704/235
7,574,362 B2 8/2009 Walker et al. 704/275
7,577,244 B2 8/2009 Taschereau 379/218.01
7,606,708 B2 10/2009 Hwang 704/257
7,620,549 B2 11/2009 Di Cristo et al. 704/257
7,634,409 B2 12/2009 Kennewick et al. 704/257
7,640,006 B2 12/2009 Portman et al. 455/412.1
7,640,160 B2 12/2009 Di Cristo et al. 704/257
7,640,272 B2 12/2009 Mahajan et al. 707/104.1
7,676,365 B2 3/2010 Hwang et al. 704/240
7,676,369 B2 3/2010 Fujimoto et al. 704/270
7,684,977 B2 3/2010 Morikawa 704/211
7,693,720 B2 4/2010 Kennewick et al. 704/275
7,729,916 B2 6/2010 Coffman et al. 704/270
7,729,918 B2 6/2010 Walker et al. 704/275
7,729,920 B2 6/2010 Chaar et al. 704/275
7,734,287 B2 6/2010 Ying 455/423
7,748,021 B2 6/2010 Obradovich 725/105
7,788,084 B2 8/2010 Brun et al. 704/7
7,792,257 B1 9/2010 Vanier
7,801,731 B2 9/2010 Odinak et al. 704/275
7,809,570 B2 10/2010 Kennewick et al. 704/257
7,818,176 B2 10/2010 Freeman et al. 704/270
7,831,426 B2 11/2010 Bennett 704/252
7,831,433 B1 11/2010 Belvin et al. 704/275
7,856,358 B2 12/2010 Ho 704/270
7,873,519 B2 1/2011 Bennett 704/257
7,873,523 B2 1/2011 Potter et al. 704/275
7,873,654 B2 1/2011 Bernard 707/769
7,881,936 B2 2/2011 Longe et al. 704/257
7,890,324 B2 2/2011 Bangalore et al. 704/231
7,894,849 B2 2/2011 Kass et al. 455/550.1
7,902,969 B2 3/2011 Obradovich 340/439
7,917,367 B2 3/2011 Di Cristo et al. 704/270.1
7,920,682 B2 4/2011 Byrne et al. 379/88.18
7,949,529 B2 5/2011 Weider et al. 704/270
7,949,537 B2 5/2011 Walker et al. 704/275
7,953,732 B2 5/2011 Frank et al. 707/724
7,974,875 B1 7/2011 Quilici et al. 705/14.4
7,983,917 B2 7/2011 Kennewick et al. 704/257
7,984,287 B2 7/2011 Gopalakrishnan et al. ... 713/100
8,005,683 B2 8/2011 Tessel et al. 704/275
8,015,006 B2 9/2011 Kennewick et al. 704/236
8,060,367 B2 11/2011 Keaveney 704/247
8,069,046 B2 11/2011 Kennewick et al. 704/257
8,073,681 B2 12/2011 Baldwin et al. 704/9
8,077,975 B2 12/2011 Ma et al. 382/187
8,082,153 B2 12/2011 Coffman et al. 704/270
8,086,463 B2 12/2011 Ativanichayaphong et al. 704/275
8,112,275 B2 2/2012 Kennewick et al. 704/240
8,140,327 B2 3/2012 Kennewick et al. 704/226
8,140,335 B2 3/2012 Kennewick et al. 704/257
8,145,489 B2 3/2012 Freeman et al. 704/257
8,150,694 B2 4/2012 Kennewick et al. 704/257
8,155,962 B2 4/2012 Kennewick et al. 704/257
8,170,867 B2 5/2012 Germain 704/9
8,195,468 B2 6/2012 Weider et al. 704/275
8,200,485 B1 6/2012 Lee et al. 704/231
8,219,399 B2 7/2012 Lutz et al. 704/270
8,219,599 B2 7/2012 Tunstall-Pedoe 707/955
8,224,652 B2 7/2012 Wang et al. 704/275
8,255,224 B2 8/2012 Singleton et al. 704/275
8,326,627 B2 12/2012 Kennewick et al. 704/257
8,326,634 B2 12/2012 Di Cristo et al. 704/270.1
8,326,637 B2 12/2012 Baldwin et al. 704/275
8,332,224 B2 12/2012 Di Cristo et al. 704/257
8,370,147 B2 2/2013 Kennewick et al. 704/257
8,447,607 B2 5/2013 Weider et al. 704/250
8,452,598 B2 5/2013 Kennewick et al. 704/257

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,509,403	B2	8/2013	Chiu et al.	379/114.13
8,515,765	B2	8/2013	Baldwin et al.	704/275
8,527,274	B2	9/2013	Freeman et al.	704/257
8,589,161	B2	11/2013	Kennewick et al.	704/252
8,620,659	B2	12/2013	Di Cristo et al.	704/257
8,719,005	B1 *	5/2014	Lee	G06F 17/271 704/10
8,719,009	B2	5/2014	Baldwin et al.	704/9
8,719,026	B2	5/2014	Kennewick et al.	704/257
8,731,929	B2	5/2014	Kennewick et al.	704/257
8,738,380	B2	5/2014	Baldwin et al.	704/257
8,849,652	B2	9/2014	Weider et al.	704/9
8,849,670	B2	9/2014	Di Cristo et al.	704/270.1
8,849,696	B2	9/2014	Pansari	
8,886,536	B2	11/2014	Freeman et al.	704/257
8,983,839	B2	3/2015	Kennewick et al.	704/254
9,015,049	B2	4/2015	Baldwin et al.	704/257
2001/0039492	A1	11/2001	Nemoto	704/231
2001/0041980	A1	11/2001	Howard et al.	704/270
2001/0049601	A1	12/2001	Kroeker et al.	704/254
2001/0054087	A1	12/2001	Flom et al.	709/218
2002/0010584	A1	1/2002	Schultz	
2002/0015500	A1	2/2002	Belt et al.	381/66
2002/0022927	A1	2/2002	Lemelson et al.	701/301
2002/0029261	A1	3/2002	Shibata	709/219
2002/0032752	A1	3/2002	Gold et al.	709/218
2002/0035501	A1	3/2002	Handel et al.	705/10
2002/0040297	A1	4/2002	Tsiao et al.	704/257
2002/0049535	A1	4/2002	Rigo et al.	701/211
2002/0049805	A1	4/2002	Yamada et al.	709/202
2002/0059068	A1	5/2002	Rose et al.	704/246
2002/0065568	A1	5/2002	Silfvast et al.	700/94
2002/0067839	A1	6/2002	Heinrich	381/110
2002/0069059	A1	6/2002	Smith	704/257
2002/0069071	A1	6/2002	Knockeart et al.	704/275
2002/0073176	A1	6/2002	Ikeda	
2002/0082911	A1	6/2002	Dunn et al.	705/14
2002/0087312	A1	7/2002	Lee et al.	704/251
2002/0087326	A1	7/2002	Lee et al.	704/270.1
2002/0087525	A1	7/2002	Abbott et al.	707/3
2002/0107694	A1	8/2002	Lerg	704/273
2002/0120609	A1	8/2002	Lang et al.	707/1
2002/0124050	A1	9/2002	Middeljans	709/203
2002/0133354	A1	9/2002	Ross et al.	704/275
2002/0133402	A1	9/2002	Faber et al.	705/14
2002/0135618	A1	9/2002	Maes et al.	345/767
2002/0138248	A1	9/2002	Corston-Oliver et al.	704/1
2002/0143532	A1	10/2002	McLean et al.	704/235
2002/0143535	A1	10/2002	Kist et al.	704/251
2002/0152260	A1	10/2002	Chen	
2002/0161646	A1	10/2002	Gailey et al.	705/14
2002/0173333	A1	11/2002	Buchholz et al.	455/527
2002/0173961	A1	11/2002	Guerra	704/258
2002/0184373	A1	12/2002	Maes	709/228
2002/0188602	A1	12/2002	Stubler et al.	707/3
2002/0198714	A1	12/2002	Zhou	704/252
2003/0014261	A1	1/2003	Kageyama	704/275
2003/0016835	A1	1/2003	Elko et al.	381/92
2003/0046346	A1	3/2003	Mumick et al.	709/205
2003/0064709	A1	4/2003	Gailey et al.	455/412
2003/0065427	A1	4/2003	Funk et al.	701/1
2003/0069734	A1	4/2003	Everhart	704/275
2003/0088421	A1	5/2003	Maes et al.	704/270.1
2003/0097249	A1	5/2003	Walker et al.	704/1
2003/0110037	A1	6/2003	Walker et al.	704/257
2003/0112267	A1	6/2003	Belrose	345/728
2003/0115062	A1	6/2003	Walker et al.	704/258
2003/0120493	A1	6/2003	Gupta	704/270.1
2003/0135488	A1	7/2003	Amir et al.	707/3
2003/0144846	A1	7/2003	Denenberg et al.	704/277
2003/0158731	A1	8/2003	Falcon et al.	704/231
2003/0161448	A1	8/2003	Parolkar et al.	379/88.17
2003/0182132	A1	9/2003	Niemoeller	704/275
2003/0204492	A1	10/2003	Wolf et al.	707/3
2003/0206640	A1	11/2003	Malvar et al.	381/93
2003/0212550	A1	11/2003	Ubale	704/215
2003/0212558	A1	11/2003	Matula	704/260
2003/0212562	A1	11/2003	Patel et al.	704/275
2003/0225825	A1	12/2003	Healey et al.	709/203
2003/0236664	A1	12/2003	Sharma	704/251
2004/0006475	A1	1/2004	Ehlen et al.	704/270.1
2004/0010358	A1	1/2004	Oesterling et al.	701/49
2004/0025115	A1	2/2004	Sienel et al.	715/513
2004/0030741	A1	2/2004	Wolton	
2004/0044516	A1	3/2004	Kennewick et al.	704/5
2004/0098245	A1	5/2004	Walker et al.	704/1
2004/0117179	A1	6/2004	Balasuriya	704/231
2004/0117804	A1	6/2004	Scahill et al.	719/320
2004/0122674	A1	6/2004	Bangalore et al.	704/276
2004/0140989	A1	7/2004	Papageorge	345/700
2004/0158555	A1	8/2004	Seedman et al.	707/3
2004/0166832	A1	8/2004	Portman et al.	455/412.1
2004/0167771	A1	8/2004	Duan et al.	704/10
2004/0172247	A1	9/2004	Yoon	
2004/0172258	A1	9/2004	Dominach et al.	704/277
2004/0193408	A1	9/2004	Hunt	704/209
2004/0193420	A1	9/2004	Kennewick et al.	704/257
2004/0199375	A1	10/2004	Ehsani et al.	704/4
2004/0205671	A1	10/2004	Sukehiro et al.	715/532
2004/0243417	A9	12/2004	Pitts, III et al.	704/276
2005/0015256	A1	1/2005	Kargman	704/272
2005/0021331	A1	1/2005	Huang et al.	704/231
2005/0021334	A1	1/2005	Iwahashi	704/240
2005/0021470	A1	1/2005	Martin et al.	705/51
2005/0021826	A1	1/2005	Kumar	709/232
2005/0033574	A1	2/2005	Kim et al.	704/251
2005/0033582	A1	2/2005	Gadd et al.	704/277
2005/0043940	A1	2/2005	Elder	704/9
2005/0080632	A1	4/2005	Endo et al.	704/277
2005/0114116	A1	5/2005	Fiedler	704/201
2005/0125232	A1	6/2005	Gadd	704/270.1
2005/0131673	A1	6/2005	Koizumi et al.	704/2
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
2005/0144013	A1	6/2005	Fujimoto et al.	704/277
2005/0216254	A1	9/2005	Gupta et al.	704/9
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
2006/0072738	A1	4/2006	Louis et al.	379/265.02
2006/0074671	A1	4/2006	Farmaner et al.	704/257
2006/0100851	A1	5/2006	Schonebeck	
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	Francois et al.	704/226
2007/0038436	A1	2/2007	Cristo et al.	704/9
2007/0038445	A1	2/2007	Helbing et al.	704/235
2007/0043569	A1	2/2007	Potter, III et al.	704/270
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
2007/0061067	A1	3/2007	Zeinstra et al.	701/200
2007/0061735	A1	3/2007	Hoffberg et al.	715/744
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
2007/0078814	A1	4/2007	Flowers	
2007/0094003	A1	4/2007	Huang et al.	704/3
2007/0112630	A1	5/2007	Lau	
2007/0118357	A1	5/2007	Kasravi et al.	704/10
2007/0124057	A1	5/2007	Prieto	
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
2007/0174258	A1	7/2007	Jones	
2007/0179778	A1	8/2007	Gong et al.	704/9
2007/0185859	A1	8/2007	Flowers	
2007/0186165	A1	8/2007	Maislos et al.	715/728

(56) **References Cited**

U.S. PATENT DOCUMENTS

2007/0192309	A1	8/2007	Fischer	
2007/0198267	A1	8/2007	Jones et al.	704/257
2007/0203736	A1	8/2007	Ashton	705/1
2007/0208732	A1	9/2007	Flowers	
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
2008/0065389	A1	3/2008	Cross	
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
2008/0154604	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
2008/0189110	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
2009/0144131	A1	6/2009	Chiu	
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
2009/0259646	A1	10/2009	Fujita et al.	707/5
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
2009/0307031	A1	12/2009	Winkler et al.	705/7
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
2011/0112921	A1	5/2011	Kennewick et al.	705/26.1
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
2011/0307167	A1	12/2011	Taschereau	
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
2013/0054228	A1	2/2013	Baldwin et al.	704/9
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
2013/0297293	A1	11/2013	Di Cristo et al.	704/9
2013/0304473	A1	11/2013	Baldwin et al.	704/257
2013/0339022	A1	12/2013	Baldwin et al.	704/257
2014/0012577	A1	1/2014	Freeman et al.	704/249
2014/0108013	A1	4/2014	Di Cristo et al.	704/254
2014/0156278	A1	6/2014	Kennewick et al.	704/254
2014/0249821	A1	9/2014	Kennewick et al.	704/257
2014/0249822	A1	9/2014	Baldwin et al.	704/257
2014/0288934	A1	9/2014	Kennewick et al.	704/254
2014/0365222	A1	12/2014	Weider et al.	704/257
2015/0019217	A1	1/2015	Di Cristo et al.	704/235
2015/0073910	A1	3/2015	Kennewick et al.	705/14.58
2015/0095159	A1	4/2015	Kennewick et al.	705/14.58
2015/0348544	A1	12/2015	Baldwin	
2015/0364133	A1	12/2015	Freeman	

FOREIGN PATENT DOCUMENTS

JP	2006-146881	6/2006
JP	2008-027454	2/2008
JP	2008-139928	6/2008
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.

Cheyer 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.

(56) **References Cited**

OTHER PUBLICATIONS

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.

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.

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.

* cited by examiner

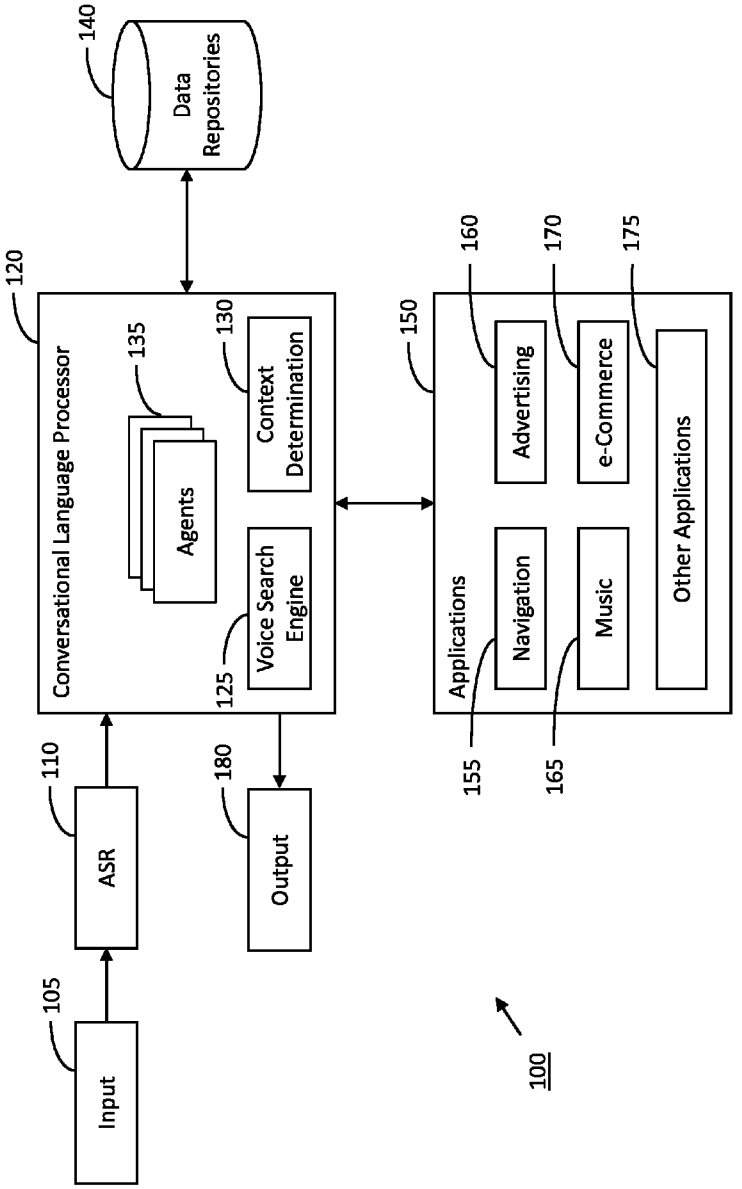

140
Data Repositories
120
Conversational Language Processor
135
Agents
130
Context Determination
125
Voice Search Engine
150
160
Advertising
170
e-Commerce
175
Other Applications
Applications
155
Navigation
165
Music
110
ASR
180
Output
105
Input
100

Figure 1

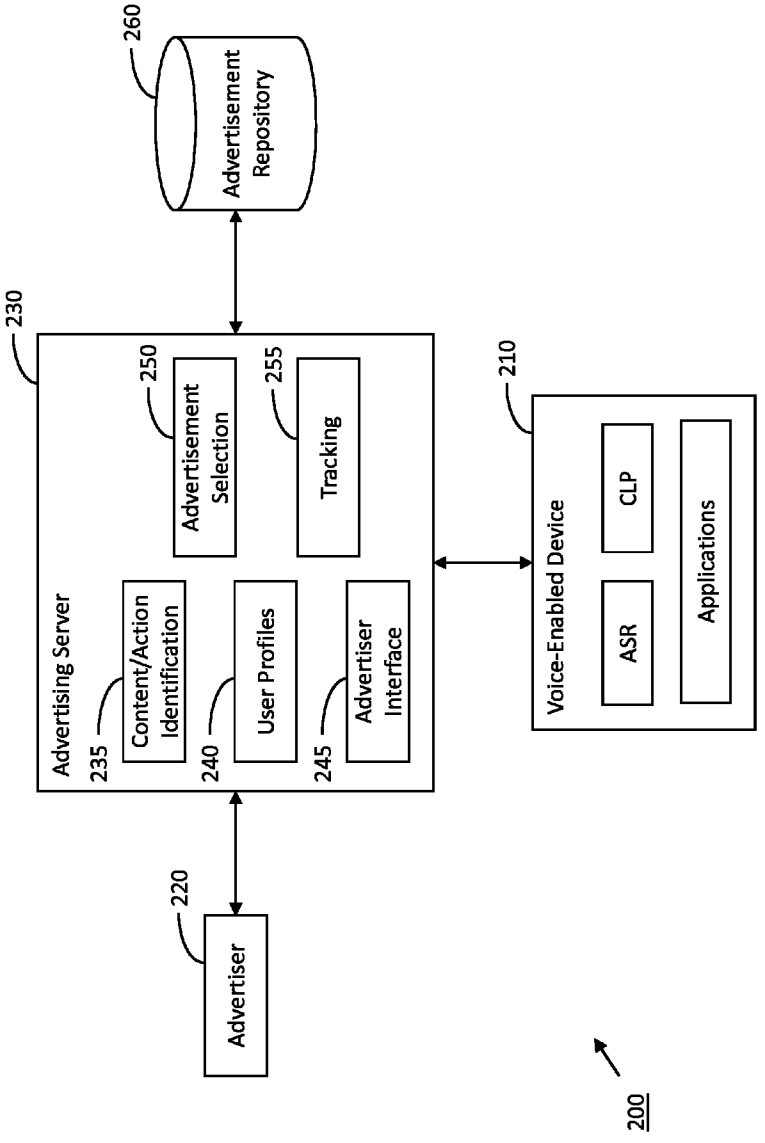

260
Advertisement Repository
230
Advertising Server
250
Advertisement Selection
255
Tracking
235
Content/Action Identification
240
User Profiles
245
Advertiser Interface
210
Voice-Enabled Device
CLP
ASR
Applications
220
Advertiser
200

Figure 2

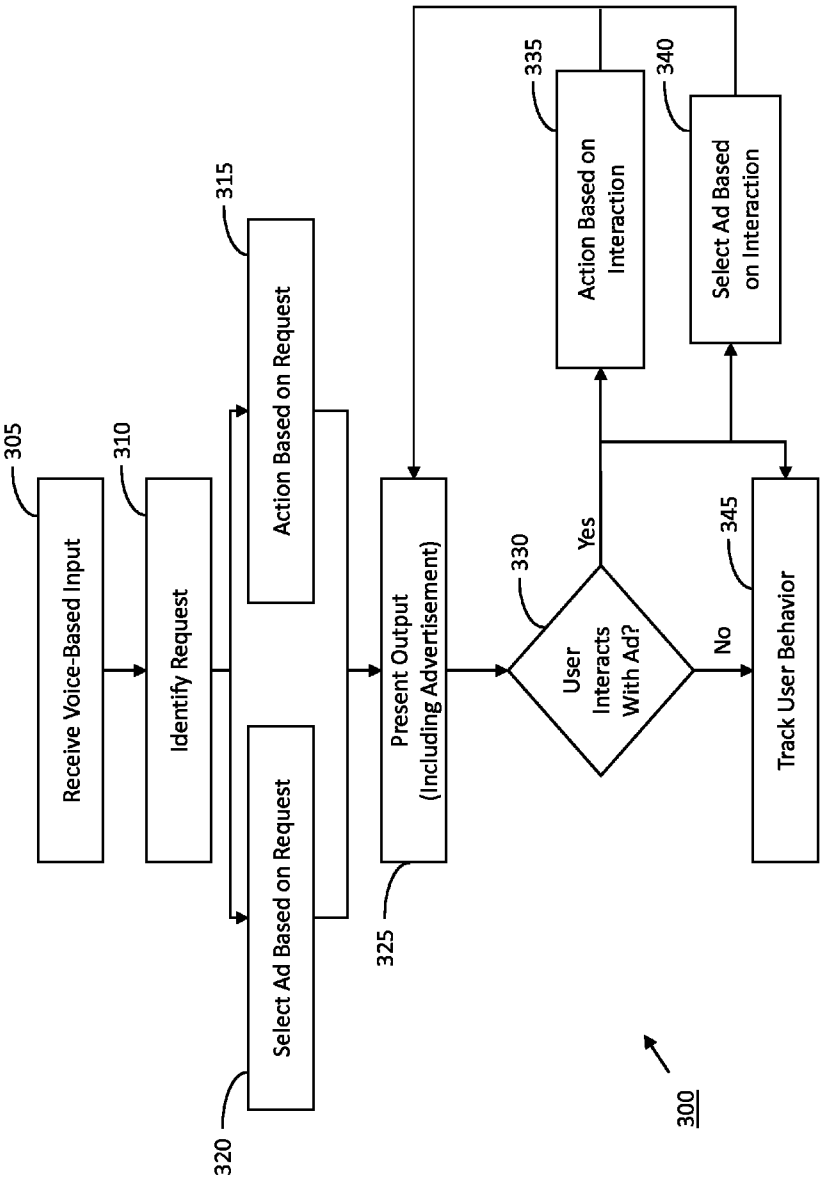

Receive Voice-Based Input
305
Identify Request
310
Action Based on Request
315
Select Ad Based on Request
320
Present Output (Including Advertisement)
325
User Interacts With Ad?
330
Yes
No
Action Based on Interaction
335
Select Ad Based on Interaction
340
Track User Behavior
345
300

Figure 3

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. 14/016,757, entitled "System and Method for Delivering Targeted Advertisements and Tracking Advertisement Interactions in Voice Recognition Context," filed Sep. 3, 2013, 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 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 communication, as would be apparent to those skilled in the art. Advertisers may specify various parameters to associate with the advertisements, such as various contexts or topic concepts (e.g., semantic indicators for a "music" concept may include words such as "music," "tunes," "songs," etc.), target demographics (e.g., a preferred audience), marketing criteria or prices for insertion (e.g., dynamic or static pricing based on various marketing criteria), or other information, as would be apparent. The advertisement repository may be associated with a server, where in response to a voice-based input from a user (e.g., at a voice-enabled device), a communications link may be established with the server. Information may be extracted from the voice-based input (e.g., words in the input, applications requested by the input, etc.), and the extracted information may be correlated with user profiles, advertisement parameters, or other information to determine which advertisements to select in relation to the voice-based input. The server may subsequently communicate the selected

advertisements to the user, and the server may track the user's subsequent interaction with the selected advertisements.

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 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 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 downloads, coupons, or any other suitable marketing, advertising, campaign, or other information, as would be apparent to those skilled in the art. A voice-enabled device **210** may receive a voice-based input and establish communications with advertising server **230**. Subsequently, advertising server **230** may select one or more advertisements from among the advertisements stored in data repository **260**, and the selected advertisements may be provided to the voice-enabled device for presentation to a user.

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, 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 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 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 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 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 res-

taurant (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.

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 providing natural language processing based on advertisements, the method being implemented on a computer system having one or more physical processors executing computer program instructions which, when executed, perform the method, the method comprising:

providing, by the computer system, an advertisement associated with a product or service for presentation to a user;

receiving, at the computer system, a natural language utterance of the user; and

interpreting, by the computer system, the natural language utterance based on the advertisement and, responsive to the existence of a pronoun in the natural language utterance, determining whether the pronoun refers to one or more of the product or service or a provider of the product or service.

2. The method of claim **1**, further comprising:

generating, by the computer system, a request related to the product or service responsive to determining that the pronoun refers to the product or service.

3. The method of claim **1**, further comprising:

generating, by the computer system, a request related to the provider of the product or service responsive to determining that the pronoun refers to the provider.

4. The method of claim **1**, wherein the natural language utterance includes a specified action and the pronoun, the method further comprising:

generating, by the computer system, a request to take the specified action with respect to the product or service responsive to determining that the pronoun refers to the product or service.

5. The method of claim **1**, wherein the natural language utterance includes a specified action and the pronoun, the method further comprising:

generating, by the computer system, a request to take the specified action with respect to the provider of the product or service responsive to determining that the pronoun refers to the provider.

6. The method of claim **1**, wherein the advertisement includes one or more of a call to action, a purchase opportunity, or a trial opportunity.

7. The method of claim **1**, further comprising:

receiving, at the computer system, another natural language utterance of the user prior to the advertisement being provided for presentation to the user; and

performing, by the computer system, speech recognition to recognize one or more words of the other natural language utterance,

wherein providing the advertisement includes providing the advertisement based on the one or more recognized words.

8. The method of claim **1**, further comprising:

receiving, at the computer system, another natural language utterance of the user prior to the advertisement being provided for presentation to the user;

interpreting, by the computer system, the other natural language utterance; and

reinterpreting and/or confirming the interpretation of the other natural language utterance based on an interaction of the user with the advertisement.

9. The method of claim **1**, further comprising:

receiving, at the computer system, another natural language utterance prior to the advertisement being provided for presentation to the user; and

determining, by the computer system, one or more requests associated with the other natural language utterance, and

wherein providing the advertisement includes providing the advertisement based on the one or more requests.

10. The method of claim **9**, wherein the one or more requests include a first request to be processed by a first device associated with the user and a second request to be processed by a second device associated with the user, and

wherein providing the advertisement includes providing the advertisement based on one or more of the first request or the second request.

11. The method of claim **10**, wherein a first device type of the first device includes one or more of a mobile phone, a navigation device, or a medial player device, and wherein the second device is of a second device type different than the first device type.

12. The method of claim **9**, wherein the one or more requests include a first request associated with a first application and a second request associated with a second application different than the first application, and

wherein providing the advertisement includes providing the advertisement based on one or more of the first request or the second request.

13. The method of **12**, wherein a first application type of the first application includes one or more of a navigation application, a music application, a commerce application, or a calling application, and wherein the second application is of an application type different than the first application type.

14. The method of claim **9**, further comprising:

obtaining, by the computer system, user profile information associated with the user, wherein the user profile information specifies prior user interactions with items;

determining by the computer system, one or more applications for processing the one or more requests, and

determining, by the computer system, categories of items based on the prior user interactions specified by the user profile information, wherein the categories relate to the one or more applications, and

wherein providing the advertisement includes providing the advertisement based on one or more of more of the categories.

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

obtaining, by the computer system, user profile information associated with the user, wherein the user profile information specifies prior user interactions with items; and

identifying, by the computer system, categories of items based on the prior user interactions specified by the user profile information, and

wherein providing the advertisement includes providing the advertisement based on one or more of the categories.

16. The method of claim **1**, further comprising:

receiving, at the computer system, a first natural language utterance during an interactive session between the user and the computer system;

interpreting, by the computer system, the first natural language utterance;

receiving, at the computer system, a second natural language utterance during the interactive session; and

reinterpreting, by the computer system, the first natural language utterance based on the second natural language utterance,

wherein providing the advertisement includes providing the advertisement based on one or more of the interpretation or the reinterpretation of the first natural language utterance.

17. The method of claim **16**, wherein providing the advertisement comprises providing the advertisement based on the reinterpretation of the first natural language utterance.

18. The method of claim **16**, further comprising:

performing, by the computer system, speech recognition to recognize one or more words of the first natural language user utterance,

wherein interpreting the first natural language utterance includes processing, based on the first natural language utterance, the one or more recognized words to determine the interpretation of the first natural language utterance, and

wherein reinterpreting the first natural language utterance comprises processing, based on the second natural language utterance, the one or more recognized words to determine the reinterpretation of the first natural language utterance.

19. The method of claim **1**, further comprising:

facilitating, by the computer system, delivery of the product or service responsive to determining that the pronoun refers to one or more of the product or service or the provider of the product or service.

20. The method of claim **19**, wherein facilitating delivery of the product or service includes performing a purchase transaction for the product or service.

21. The method of claim **19**, wherein the product or service comprises a particular product, and wherein facilitating delivery of the product or service includes providing the particular product to a user device of the user.

22. The method of claim **1**, wherein the product or service includes one or more of an application, a ringtone, a song, a video, a coupon, or a product related to a scheduled event.

23. A system providing natural language processing based on advertisements, 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 an advertisement associated with a product or service for presentation to a user;

receive a natural language utterance of the user; and

interpret the natural language utterance based on the advertisement and, responsive to the existence of a pronoun in the natural language utterance, determine whether the pronoun refers to one or more of the product or service or a provider of the product or service.

* * * * *

EXHIBIT N

USPN 9,269,097 ("the '097 patent")
Infringement Chart for Claim 1

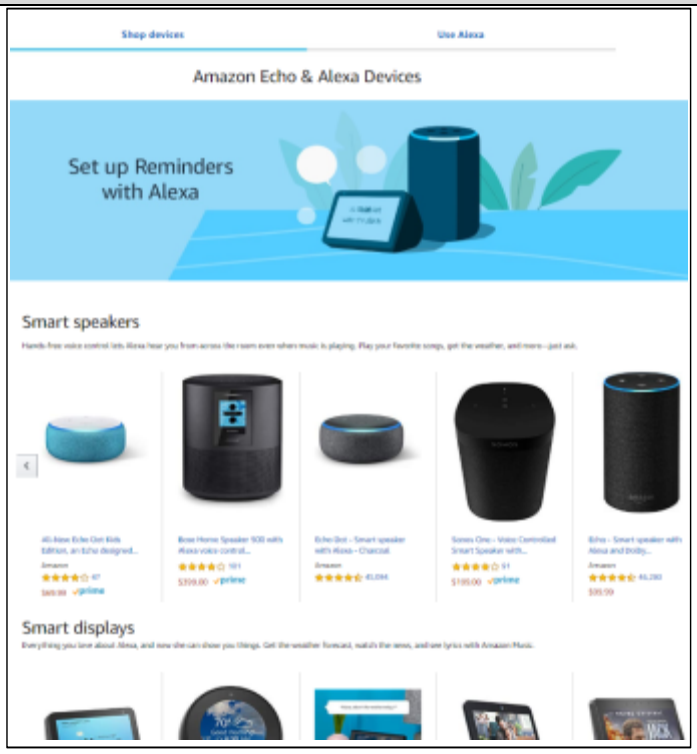
The following chart provides exemplary infringement allegations with respect to an exemplary claim of the '097 patent based on Plaintiff's investigation to date. Plaintiff reserves the right to provide modified and/or more detailed allegations (including with additional claims, products, and evidence) in accordance with its discovery obligations in view of any applicable scheduling order. This chart uses "Alexa Products" to refer to the full scope of instrumentalities identified in the complaint. Any discussion of specific Alexa Products in the following chart is exemplary.

Exemplary information relevant to infringement:

1. *Amazon Echo & Alexa Devices* - Amazon Official Site, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/?ie=UTF8&node=9818047011&ref_=sv_devicesubnav_1 (last visited Jul. 22, 2019)
2. *Alexa and Alexa Device FAQs*, Amazon Help & Customer Service - Amazon Official Site, https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (last visited Jul. 22, 2019)
3. Chaim Gartenberg, *How to shop on Amazon using Alexa voice controls*, The Verge, Jul. 10, 2017, https://www.theverge.com/2017/7/10/15947672/amazon-alexa-voice-controls-shopping-prime-echo-how-to ("*How to Shop*")
4. *Audio Hardware Configurations*, Alexa Voice Service, developer.amazon.com, https://developer.amazon.com/docs/alexa-voice-service/audio-hardware-configurations.html (last visited Jul. 22, 2019)
5. *Designing for Conversation – Build for dialog*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/ask-resources/guided/conversational-design-workshop#/lessons/gDcQVF9nOFOtNljrELzARIGN4HuD7LNh?_k=q4qp1a (last visited Jul. 22, 2019)
6. *Amazon Echo Teardown*, IFIXIT, Dec. 16, 2014, https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953
7. Richard Baguley & Colin McDonald, *Appliance Science: Alexa, how does Alexa work? The science of the Amazon Echo*, CNET, Aug. 4, 2016, https://www.cnet.com/news/appliance-science-alexa-how-does-alexa-work-the-science-of-amazons-echo/ ("*Appliance Science*")
8. Taylor Martin, *How to shop with Alexa on Prime Day*, CNET, Jul. 15, 2018, https://www.cnet.com/how-to/how-to-shop-with-alexa-prime-day/ ("*Prime Day*")
9. *What is Conversational AI?*, Amazon Alexa - Amazon Developer Services and Technologies - Amazon Official Site, https://developer.amazon.com/alexa-skills-kit/conversational-ai (last visited Jul. 22, 2019)

'097 Patent Claim Language	**Exemplary Alexa Products**
1.[pr] A method for providing natural language processing based on advertisements, the method being implemented on a computer system having one or more physical processors executing computer program instructions which, when executed, perform the method, the method comprising:	Alexa Products are programmed to perform and have performed methods for providing natural language processing based on advertisements, the method being implemented on a computer system having one or more physical processors executing computer program instructions which, when executed, perform the recited method. For example, Alexa Products (e.g. Echo, Echo Dot, Alexa apps, and Alexa cloud) include Alexa.

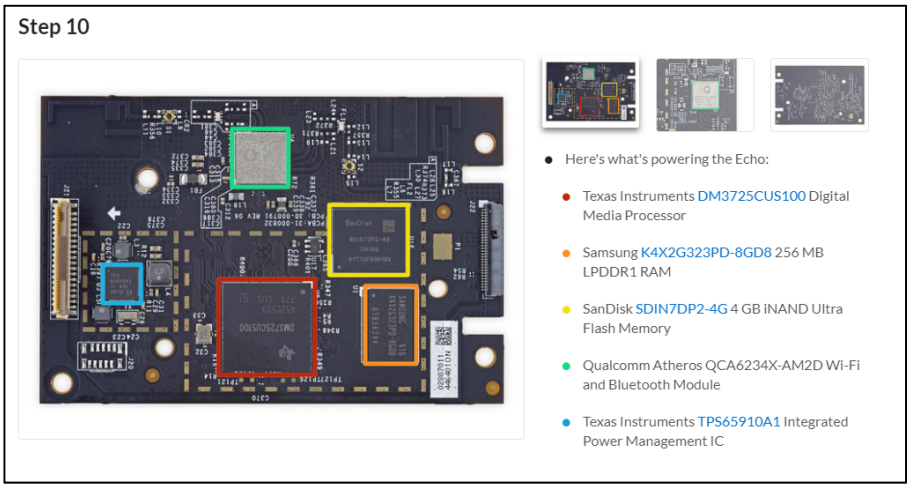
'097 Patent Claim Language	Exemplary Alexa Products



E.g. *Amazon Echo & Alexa Devices*.

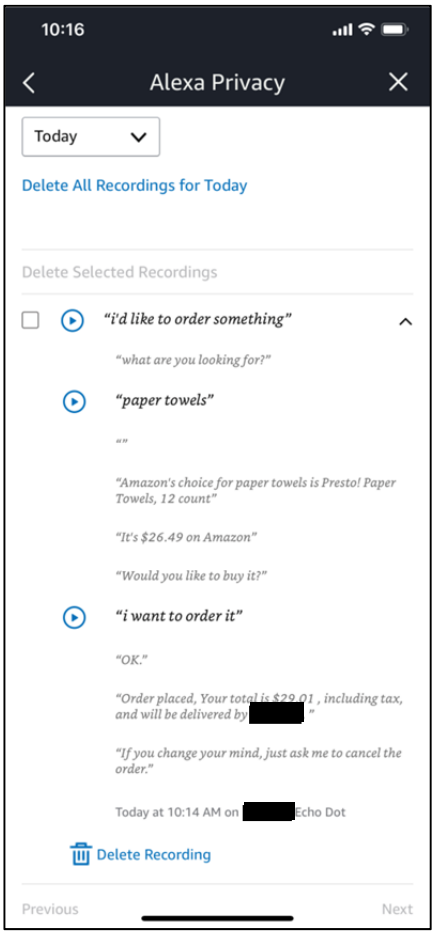
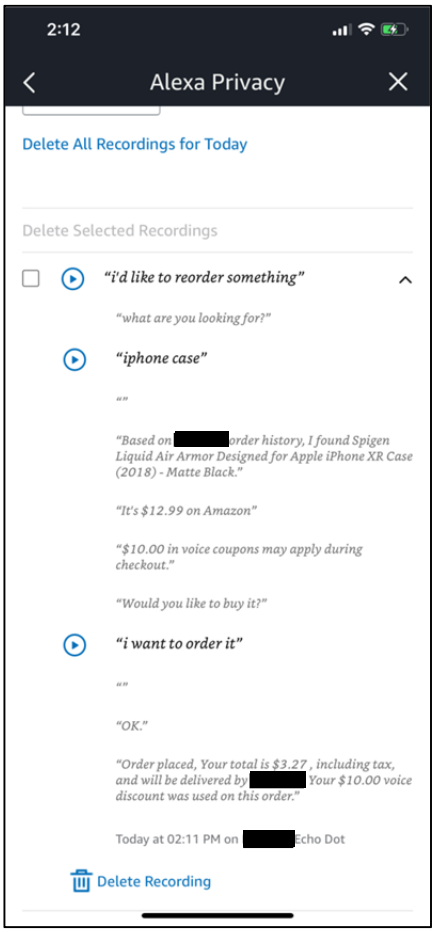
"With conversational AI, voice-enabled devices like Amazon Echo . . . ." E.g. *What is Conversational AI?*

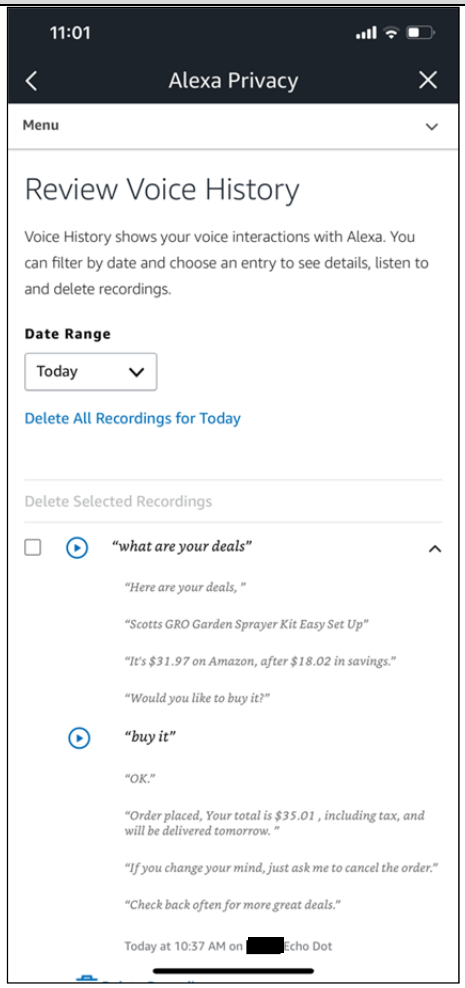
For example, the Echo has the physical processors shown below:

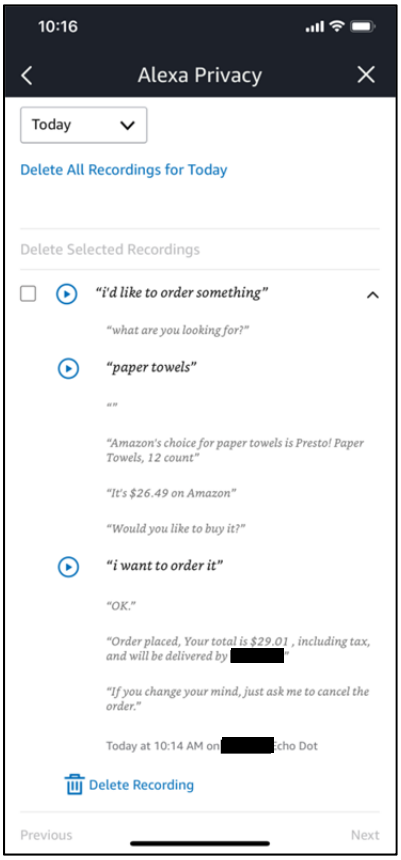
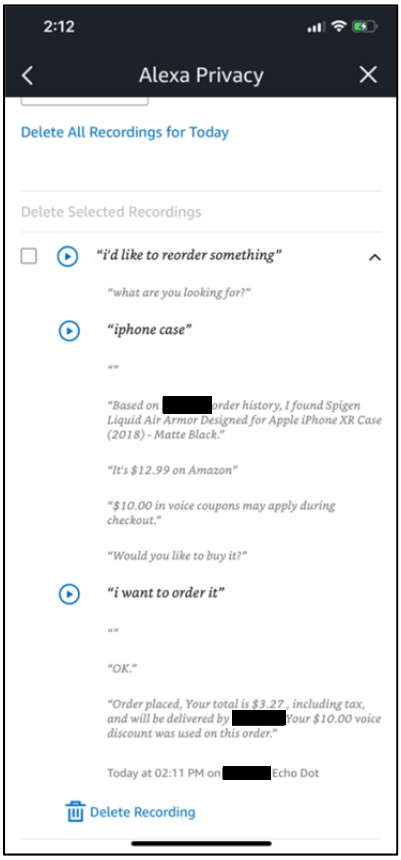


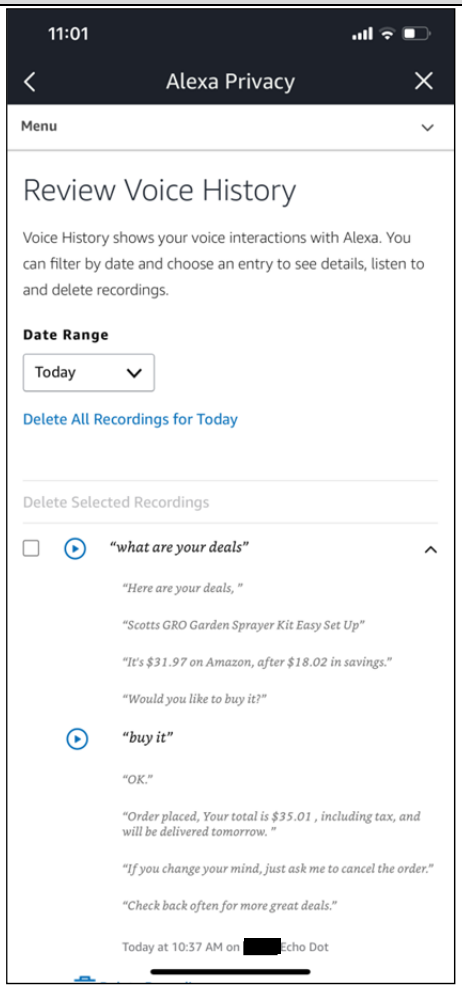
E.g. *Amazon Echo Teardown* at Step 10.

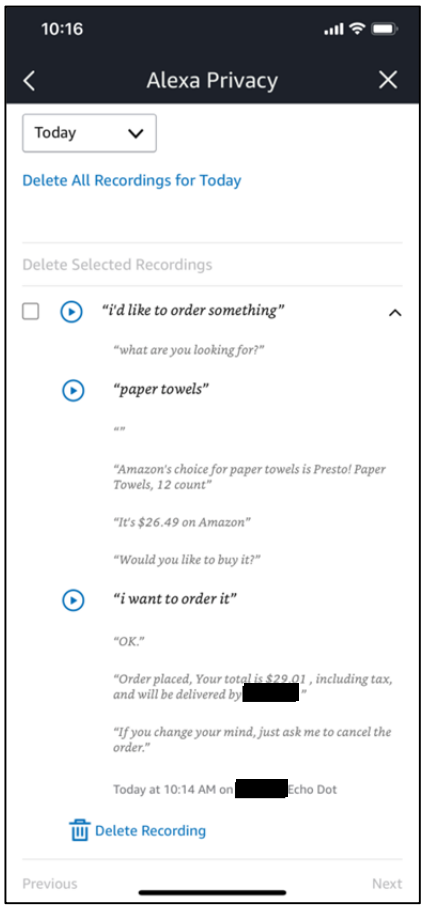
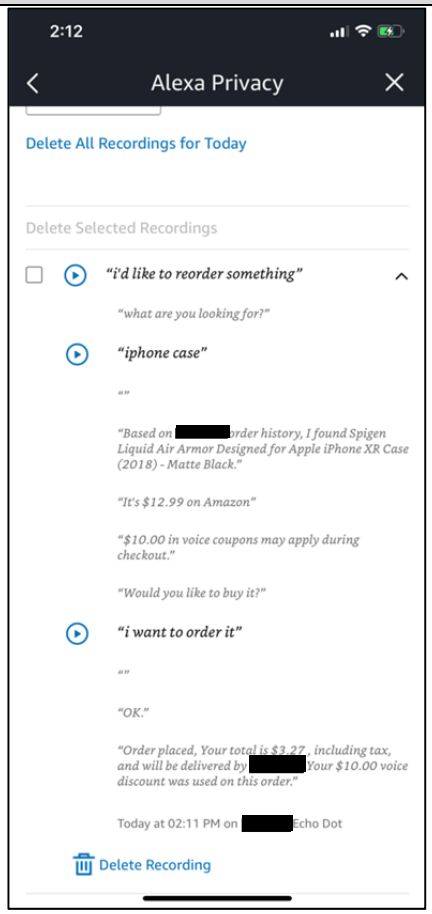
As another example, Alexa Products include Amazon's Alexa cloud that has one or more physical processors: "Alexa allows you to ask questions and make requests using just your voice. For instance, you can ask Alexa a question, such as "What is the weather today in New York?", or tell Alexa to do something, like "Add batteries to my shopping list," or "Play jazz." When you speak to Alexa, a recording of what you asked Alexa is sent to Amazon's cloud so we can process and respond to your request." E.g. *Alexa and Alexa Device FAQs* at answer to question 1.

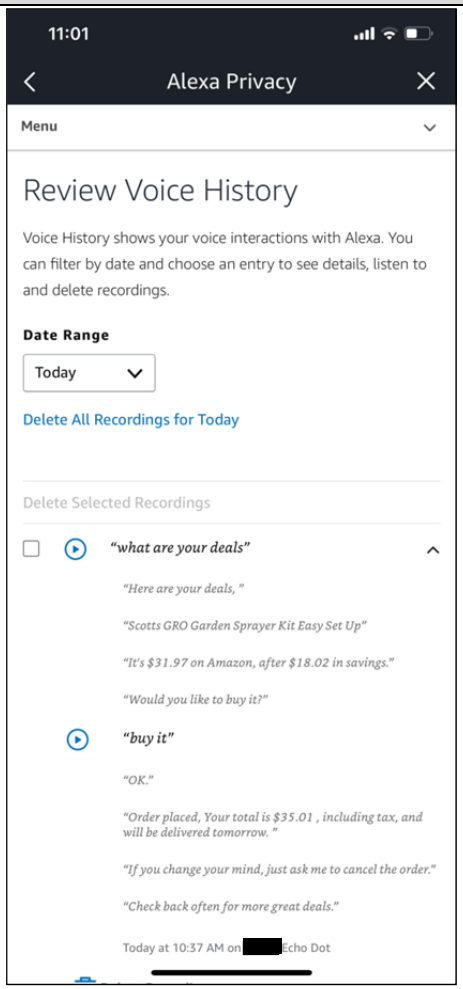
'097 Patent Claim Language	Exemplary Alexa Products
	When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*. "Alexa uses a combination of automatic speech recognition (ASR) and natural language understanding (NLU) to understand user speech and respond with precision. ASR converts user speech into text, NLU converts that text into intents for Alexa to act on. At the end of this process Alexa sends directives to your product instructing it to perform an action..." *Audio Hardware Configurations* at Automatic Speech Recognition Profiles.
[1.a] providing, by the computer system, an advertisement associated with a product or service for presentation to a user;	Alexa Products provide, by the computer system, an advertisement associated with a product or service for presentation to a user. Examples of providing an advertisement associated with a product or service are shown in screenshots of Alexa app's "Review Voice History" page below.  

'097 Patent Claim Language	Exemplary Alexa Products
	 Other examples are below. When ordering something not in your existing order history, Alexa "will suggest an 'Amazon Choice' product that meets the description." *How to Shop*. "For starters, if you want to reorder something you've already purchased from Amazon before, just say, "Alexa, reorder a Swiss Army knife." Alexa will pull up your order history and let you select from the items. You confirm with Alexa which item you want from your history and you're done." *Prime Day*. "To order something you've not ordered through Amazon before, say something like "Alexa, order deodorant," or "Alexa, buy a microSD card." Alexa will then perform a search for that item and offer a few recommendations based on Amazon preferred products or products you've purchased from similar categories in the past. It will talk you through the handful of recommendations. Like with reordering, just tell Alexa the number of the suggestion you want to add it to your cart." *Prime Day*.
[1.b] receiving, at the computer system, a natural language utterance of the user; and	Alexa Products receive, at the computer system, a natural language utterance of the user. Examples of receiving utterances are shown below in screenshots of

'097 Patent Claim Language	Exemplary Alexa Products
	Alexa app's "Review Voice History" page below.  

'097 Patent Claim Language	Exemplary Alexa Products
	 Another example is: "Each line of the script spoken by the user is known as an utterance. The utterances are then mapped to an intent. When the user speaks one of them the associated intent will be invoked." *Designing for Conversation – Build for dialog* at Step 2.
[1.c] interpreting, by the computer system, the natural language utterance based on the advertisement and, responsive to the existence of a pronoun in the natural language utterance, determining whether the pronoun refers to one or more of the product or service or a provider of the product or service.	Alexa Products interpret, by the computer system, the natural language utterance based on the advertisement and, responsive to the existence of a pronoun in the natural language utterance, determining whether the pronoun refers to one or more of the product or service or a provider of the product or service. Examples are interpreting the natural language utterances, as shown in screenshots of Alexa app's "Review Voice History" page below.

’097 Patent Claim Language	Exemplary Alexa Products
	 

'097 Patent Claim Language	Exemplary Alexa Products
	 Another example is: "The service that processes this recording is called Alexa Voice Services (AVS). Run by Amazon, this converts the recording into commands that it interprets. It's more than a simple voice-to-text service: it is a fully programmable service that can work with other online services to do a surprising range of things." *Appliance Science*.