

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM SERVICES LLC,

Petitioner,

v.

VB ASSETS, LLC,

Patent Owner.

Case: IPR2025-01379

U.S. Patent No. 10,755,699

DECLARATION OF PADHRAIC SMYTH, PH.D.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	8
II. BACKGROUND AND QUALIFICATIONS	8
III. COMPENSATION AND RELATIONSHIP WITH PARTIES.....	12
IV. MATERIALS CONSIDERED	12
V. BASIS OF OPINION	15
A. Legal Standard for Anticipation	15
B. Legal Standard for Obviousness	15
C. Level of Ordinary Skill in the Art.....	17
D. Claim Construction.....	18
VI. OVERVIEW OF THE '699 PATENT	19
A. Filing and Specification	19
B. Prosecution History	22
VII. RELATED LITIGATION AND RELATED PATENTS	24
VIII. PRIOR ART TO THE CHALLENGED PATENT	26
A. Technological Background at Time of Filing.....	26
B. Summary of the Prior Art.....	32
1. Kennewick (EX1004)	32
2. Cooper (EX1005).....	35
3. Matsuda (EX1006; EX1007)	36
IX. INVALIDITY	38
A. Claims 1-11 are Rendered Obvious by Kennewick, Cooper, and Matsuda	39

TABLE OF CONTENTS (Continued)

	Page
1. Claim 1	39
[1.pre] “A computer-implemented method of generating natural language system responses adapted based on a user's manner of speaking, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:”	39
[1.a]: “receiving, by the computer system, a user input comprising a natural language utterance;”	49
[1.b] “recognizing, by the computer system, one or more words or phrases from the natural language utterance;”	50
[1.c] “identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;”	52
[1.d] “determining, by the computer system, an interpretation of the natural language utterance based on the identified context;”	54
[1.e] “accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;”	56

TABLE OF CONTENTS

(Continued)

	Page
[1.f] “accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;”	65
[1.g] “identifying, by the computer system, a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and”	73
[1.h] “generating, by the computer system, a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.”	79
2. Claim 2: “The method of claim 1, wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.”	84
3. Claim 3 “The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response based on the identified manner comprises varying one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.”	85
4. Claim 4.....	89
[4.pre] “The method of claim 1, the method further comprising:”	89
[4.a] “obtaining, by the computer system, contextual signifiers and/or grammatical rules,”	89

TABLE OF CONTENTS (Continued)

	Page
[4.b] “wherein generating the response based on the identified manner in which the natural language utterance was spoken comprises using the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.”.....	92
5. Claim 5.....	94
[5.pre] “The method of claim 1, wherein the long-term knowledge is associated with a first user, the method further comprising:”	94
[5.a] “generating, by the computer system, a profile associated with the first user based on the long-term knowledge, wherein the context for the natural language utterance is determined based further on the profile associated with the first user.”	95
6. Claim 6.....	96
[6.pre] “The method of claim 1, wherein generating the response based on the identified manner comprises:”	96
[6.a] “adapting, by the computer system, the response based on a response format associated with the identified manner.”	96
7. Claim 7.....	98
[7.pre] “The method of claim 1, wherein recognizing the one or more words or phrases from the natural language utterance comprises:”	98
[7.a] “providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and”	99

TABLE OF CONTENTS (Continued)

	Page
[7.b] “obtaining, by the computer system, the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.”.....	100
8. Claim 8.....	101
[8.pre] “The method of claim 1, the method further comprising:”	101
[8.a] “causing, by the computer system, the response to the natural language utterance to be provided to the user.”	101
9. Claim 9: “The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response to the natural language utterance based on the identified manner comprises varying one or more of word use and/or jargon of the voice response.”	102
10. Claim 10.....	103
[10.pre] “The method of claim 1, further comprising:”	103
[10.a] “adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.”.....	103
11. Claim 11: “The method of claim 1, the method further comprising generating a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.”.....	106
B. Claims 12-22 are Rendered Obvious by Kennewick, Cooper, and Matsuda	110

TABLE OF CONTENTS
(Continued)

	Page
1. Claim 12	110
[12.pre] “A system for generating natural language system responses adapted based on a user's manner of speaking, the system comprising:”	110
[12.a] “one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to:”	110
2. Claims 13-22	111
X. CONCLUSION.....	112

I, Padhraic Smyth, Ph.D., declare as follows:

I. INTRODUCTION

1. I have been asked by the party requesting this review, Amazon.com Services LLC (“Petitioner”), to provide my expert opinions in support of the petition for Inter Partes review of U.S. Patent No. 10,755,699 (the “’699 patent” or “challenged patent”) challenging the patentability of claims 1-22 (the “challenged claims”).

2. In summary, it is my opinion that the challenged claims are disclosed and rendered obvious by the prior art. I set forth below my detailed opinions.

II. BACKGROUND AND QUALIFICATIONS

3. The details of my education, work experience, research, patents, and publications are summarized in my CV, attached as EX1014. I provide here a summary of some of the relevant information from my CV, as well as other relevant background information.

4. I have a Ph.D. and Master’s degree in Electrical Engineering from the California Institute of Technology. I earned a Bachelor of Engineering (Electronic) from the National University of Ireland, University College Galway in 1984. My senior undergraduate thesis was on speech recognition.

5. At present, I am a Distinguished Professor and Endowed Chair in Artificial Intelligence in the Department of Computer Science at the University of

California, Irvine. Prior to that, I was a Member of Technical Staff and Technical Group Leader at the Jet Propulsion Laboratory.

6. As a professor at UC Irvine since 1996, I conduct research in the areas of machine learning, artificial intelligence, signal processing, pattern recognition, and natural language processing, and I regularly teach courses related to these topics. I have published over 230 peer-reviewed papers and two books on these topics and my work has been cited over 70,000 times in the research literature. My research work is funded by federal agencies such as NSF, NIH, NIST, IARPA, DARPA, ONR, and NASA.

7. During my career I have worked extensively on research related to key aspects of automated speech recognition, natural language processing, user modeling, and human dialog. As an example, since 1992 I have conducted a significant body of research and published a series of papers on the topic of hidden Markov models (HMMs). HMMs are a general statistical framework for pattern recognition in sequential data (such as audio waveforms or text) and were a key component in automated speech recognition systems and natural language understanding since the 1980's. My research on HMMs produced significant advances in our understanding of the statistical and algorithmic foundations of speech recognition and natural language understanding. These advances included development of new HMM methods for detecting anomalous patterns in time-series

signals, new algorithms for automated clustering of waveforms using HMMs, new classes of HMMs with applications to speech modeling, and novel HMM methods for automated segmentation of waveforms. My papers on the topic of HMMs have been collectively cited over 1,800 times in the literature.

8. I have also conducted extensive research in the field of natural language understanding since 1999. My work included co-authoring a textbook, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, published in 2003 with chapters devoted to algorithms for natural language understanding. I have also written papers on topics like analysis of character sequences in non-English languages, automated algorithms for extracting human-interpretable summaries of content from large text corpora, methods for automated characterization of authors from the text they generate, automated techniques for creating network and graph models of relationships between documents and entities mentioned in the documents, algorithms for analyzing text with the assistance of human-generated semantic ontologies, and analysis of large language models. As with my work on HMMs, my work in natural language understanding meant that I was, and am, broadly familiar with the research literature in this area.

9. I have also conducted research in the area of user modeling, focusing on statistical and machine learning techniques for automated generation of user models from observational data. In addition, I have published papers on techniques

for automated analysis of human dialog, such as automatically classifying utterances into specific topics. As I discuss below, user modeling is an important function of many interactive conversational systems, which enables such systems to manage a dialogue with the user and infer information about the user's requests.

10. In addition to my research and publications I have taught and lectured extensively on these topics. Since my appointment as a professor at UC Irvine in 1996, I have taught both undergraduate and graduate courses there in machine learning, pattern recognition, natural language understanding, artificial intelligence, and user modeling. In addition, I have taught professional courses on these topics at UC Irvine and have given keynote lectures and tutorials on these topics at a variety of international conferences since the late 1990s.

11. In my professional capacity I have also served in editorial roles and conference chair positions for major journals and conferences in my research areas, including being a founding editorial board member for the Journal of Machine Learning Research, an associate editor for the Journal of the American Statistical Association, program chair for the ACM SIGKDD conference on Knowledge Discovery and Data Mining, co-program chair for the International Joint Conference on Artificial Intelligence, and have served as Senior Area Chair for the NeurIPS conference and the International Machine Learning Conference on multiple occasions.

12. I am also a technical consultant with private industry in areas including machine learning, user modeling, signal processing, pattern recognition, and natural language understanding. Over the past 25 years I have consulted in these areas for companies including Toshiba, Microsoft, Google, SAP, Netflix, eBay, AT&T, Samsung, Nokia, as well as several startup companies. Accordingly, I have knowledge of and experience with the practical application, in terms of operational use and commercial products, of methods from the fields of machine learning, user modeling, signal processing, pattern recognition, and natural language understanding.

III. COMPENSATION AND RELATIONSHIP WITH PARTIES

13. I am being compensated for my time at a rate of \$800 per hour. This compensation is not contingent upon my performance, the conclusions I reach in my analysis, the outcome of this matter, or any issues involved in or related to this matter.

14. I have no financial interest in Petitioner or any related parties.

IV. MATERIALS CONSIDERED

15. I have reviewed and considered, in preparing this declaration, the '699 patent (EX1001), its prosecution history (EX1003), and the following prior art references:

- U.S. Patent App. Pub. No. 2004/0193420, published September 30, 2004 and titled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance” (“Kennewick”) (EX1004);
- U.S. Patent No. 6,757,362, issued on June 29, 2004 and titled “Personal Virtual Assistant” (“Cooper”) (EX1005);
- Japanese Laid Open Patent Publication No. JP2004334591, published November 25, 2004 and titled “Conversation System and Conversation Processing Program” (“Matsuda”) (EX1006, EX1007)

16. I am also aware of and considered in forming my opinions presented in this declaration, other information available to a person of ordinary skill in the art at the time of the effective filing date of the challenged patent (“POSITA”), such as other patents, reference materials, articles, and standards, and knowledge of a POSITA.¹ These include:

¹ I understand that the challenged patent has an effective filing date of October 16, 2006.

- The American Heritage Dictionary of the English Language, 4th Ed. (2000), Houghton Mifflin, Co., ISBN 0-395-82517-2, (EX1008);
- Asthana, A. & Krzyzanowski, P., “A Small Domain Communications System for Personalized Shopping Assistance,” Proceedings of 1994 International Conference on Personal Wireless Communications (pp. 199-203), IEEE Press (1994) (“Asthana”) (EX1009);
- Lucente, M., “Conversational interfaces for e-commerce applications,” Communications of the ACM (Sept. 2000), 43(9), 59-61, (“Lucente”) (EX1010);
- Huang et al., “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development,” Prentice Hall (2001) (“Huang”) (EX1011);
- U.S. Patent No. 6,859,776, issued February 22, 2005 and titled “Method and Apparatus for Optimizing a Spoken Dialog Between a Person and a Machine” (“Cohen”) (EX1012).

17. Should new information be disclosed or discovered during this proceeding, I may supplement and further elaborate on the opinions I express in this declaration as necessary. Nothing in this declaration should be interpreted as an

opinion that no grounds of invalidity exist other than the ones in the petition. I do not express such an opinion and have not reached such an opinion.

V. BASIS OF OPINION

A. Legal Standard for Anticipation

18. I understand that a claimed invention is unpatentable as “anticipated” by a single prior art reference if it discloses each and every limitation of the claim. I also understand that, even if a claim is not explicitly disclosed in the prior art reference, the prior art reference may nevertheless anticipate the claim if the subject matter of the claim is inherently present from the perspective of a POSITA. Moreover, I understand that even if the prior art reference does not disclose the limitations of a claim in the same arrangement or combination of the limitations of the claim, it may nevertheless anticipate if a POSITA would envisage the claimed arrangement or combination.

B. Legal Standard for Obviousness

19. I understand that a claimed invention is unpatentable as “obvious” if the differences between the invention and the prior art are such that the subject matter of the claim as a whole would have been obvious to a POSITA.

20. It is my understanding that “obviousness” is a question of law based on underlying factual issues including (1) the scope and content of the prior art, (2) the differences between the prior art and the asserted claims, (3) the level of ordinary

skill in the pertinent art, and (4) the existence of secondary considerations such as commercial success, long-felt but unresolved needs, and failure of others.

21. I understand that a single reference can render a patent claim obvious if any differences between that reference and the claims would have been obvious to a POSITA. Alternatively, the teachings of two or more references may be combined in the same way as disclosed in the claims, if such a combination would have been obvious to a POSITA. In determining whether a combination based on either a single reference or multiple references would have been obvious, I understand that it is appropriate to consider the following factors:

- Whether the teachings of the prior art references disclose known concepts combined in familiar ways, and when combined, would yield predictable results;
- Whether a POSITA could implement a predictable variation, and would see the benefit of doing so;
- Whether a POSITA would have recognized a reason to combine known elements in the manner described in the claim;
- Whether there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the patent; and

- Whether the innovation applies a known technique that had been used to improve a similar device or method in a similar way.

22. In an obviousness evaluation based on a combination of multiple prior art references, I understand that the prior art references themselves may provide a suggestion, motivation, or reason to combine, but other times the nexus linking two or more prior art references is simple common sense. I further understand that obviousness analysis recognizes that market demand, rather than scientific literature, often drives innovation, and that a motivation to combine references may be supplied by the direction of the marketplace.

C. Level of Ordinary Skill in the Art

23. I understand that the '699 patent should be interpreted from the perspective of a person having "ordinary skill in the art" as of the effective filing date of the patent application (i.e., a POSITA). For the purposes of this declaration, I assume that the effective filing date of the patent application leading to the '699 patent is no earlier than October 16, 2006, and I understand that Patent Owner has not claimed an earlier priority date in the parallel litigation. I am familiar with the technological field at issue and the state of the art as of that date.

24. I understand that the education level of those in the technological field, sophistication of the technology, types of problems encountered, prior art solutions

to those problems, and the speed at which innovations are made in the field may factor into establishing what the appropriate level of ordinary skill should be.

25. In my opinion, a person having ordinary skill in the art of the '699 patent at the time of the effective filing date would have at least a Bachelor-level degree in computer science, computer engineering, electrical engineering, or a related field in computing technology, and two years of experience with automatic speech recognition and natural language understanding, or equivalent education, research experience, or knowledge.

26. As I describe my qualifications above and as my CV illustrates, I had at least the requisite level of skill with respect to the subject matter of the '699 patent at the time of the effective filing date.

D. Claim Construction

27. I understand that, in *inter partes* review ("IPR") proceedings, claims are construed using the same claim construction standard that district courts use, which I understand is called the "*Phillips*" claim construction standard, including construing the claim in accordance with the plain and ordinary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

28. I applied plain and ordinary meaning to all terms of the challenged claims for my analysis in this proceeding.

29. I understand the district court in *VB Assets, LLC, v. Amazon.com, Inc., et al.*, C.A. No. 19-1410 (MN) (D. Del.) issued a claim construction order involving U.S. Patent Nos. 8,073,681 (“the ’681 Patent”) and 9,015,049 (“the ’049 Patent”), parents of the challenged patent that share the same specification. EX1015. The order indicates that the parties agreed the term “short-term knowledge” did not require construction and the district court found that the terms “long-term knowledge” and “context” did not require construction. EX1015, 1-2.

30. I am not an attorney and my opinions are limited to what I believe a POSITA would have understood the meaning of certain claim terms to be, based on the patent specifications and prosecution histories in light of the POSITA’s knowledge.

VI. OVERVIEW OF THE ’699 PATENT

A. Filing and Specification

31. The ’699 patent was filed as Application No. 16/417,178 (the “’178 application”) on May 20, 2019. The ’178 application is a division of Application No. 14/691,445 (the “’445 application”) filed April 20, 2015, now Patent No. 10,297,249. The ’445 application is a continuation of Application No. 13/987,645 (the “’645 application”) filed August 19, 2013, now Patent No. 9,015,049. The ’645 application is a division of Application No. 13/251,712 (the “’712 application”) filed October 3, 2011, now Patent No. 8,515,765. The ’712 application is a continuation

of Application No. 11/580,926, filed October 16, 2006 (the “effective filing date”), now Patent No. 8,073,681.

32. The ’699 patent describes an interface that receives natural language inputs in the form of utterances spoken by a user. EX1001, 2:17-21, 11:30-35. The “utterance component of the input may be processed by a speech recognition engine” to generate an interpretation of the utterance. *Id.*, 2:37-41.

33. Context is used to disambiguate utterances and improve interpretation accuracy. *Id.*, 4:32-36. For example, the word “traffic” may refer to a road condition, a music band, or a movie, depending on the context. *Id.*, 11:45-55. The context may be used within a “[c]ooperative conversation model...to implement a conversational Human-to-Machine interface that reflects how humans interact with each other and their normal behavior in conversation.” *Id.*, 13:51-56.

34. An Intelligent Hypothesis Builder “generate[s] one or more explicit hypotheses of a user’s intent when an utterance contains all information...needed to complete a request or task.” *Id.*, 16:43-47. It uses “short-term shared knowledge...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.” *Id.*, 16:24-31. “Short-term knowledge may accumulate during a single conversation, where input received during a single conversation may be retained.” *Id.*, 5:10-12. The long-term

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.” *Id.*, 5:29-33.

35. Both short-term and long-term knowledge about the user may be used to generate a response to the user. *Id.*, 14:3-5. The “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.” *Id.*, 14:64-15:3. An Adaptive Response Builder uses the output of the Intelligent Hypothesis Builder “to formulate [a] personalized and conversational response” to the user. *Id.*, 16:63-65, 17:5-8. The response “may include a verbal or audible reply played through an output device..., and/or an action performed by a device, computer, or machine.” *Id.*, 18:18-23.

36. The response is further based on the user’s manner of speaking. *Id.*, 18:46-47. “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).” *Id.*, 18:45-50. The system identifies a user’s manner of speaking based on accumulated short-

term and long-term knowledge. *Id.*, 18:54-58. The examples at column 19, lines 1-19 of the '699 patent “demonstrate how a response may adapt to a user’s input word choices and manner of speaking.” *Id.*, 18:65-67.

B. Prosecution History

37. I reviewed the prosecution history of the '699 patent. EX1003. The Applicant filed multiple information disclosure statements (“IDS’s”) collectively citing more than one thousand references when the application that resulted in the '699 patent was initially filed. EX1003, 416-421, 429-489, and 493-528. One of these IDS’s listed the Kennewick reference (EX1004) asserted in this petition (U.S. Patent Application Publication No. 2004/0193420). *Id.*, 434. These IDS’s also listed many other references naming Kennewick as an inventor. *Id.*, 417-419, 432, 440, 443, 446-454, 456, 471-473, 475-482, and 523-525. After the application that resulted in the '699 patent was filed, during prosecution of the application, the Applicant filed additional IDS’s listing several more references (including additional references naming Kennewick as an inventor). *Id.*, 404-405, 373, 171, and 27-28.

38. The Examiner issued an Office Action dated December 18, 2019 rejecting the independent claims as being obvious based on a different Kennewick reference, U.S. Patent App. Pub. No. 2004/0044516 (“Kennewick516”) in combination with U.S. Patent App. Pub. No. 2003/0167167 (“Gong”). *Id.*, 211-254. The Examiner also noted in the Office Action that originally-filed claim 6, which

ultimately issued as limitations [1.f] and (in part) [1.g] of claim 1 of the '699 patent, recites allowable subject matter. *Id.*, 248. The Examiner cited Perkins (US 7,072,888) as disclosing intervening claim 5's recitation of "accumulating ... short-term knowledge based on one or more natural language utterances received during a predetermined time period." *Id.*, 246-248, 574.

39. In indicating the features of claim 6 (which recites "accumulating ... long-term knowledge ... based on one or more natural language utterances received prior to the predetermined time period") are allowable, the Examiner noted that Perkins "teaches where 'short-term' knowledge includes queries within the last n days, but describes where data older than n days is obsolete and is deleted (therefore not obvious to combine with an additional reference that teaches long-term data based on utterances received more than n days ago)." *Id.*, 248, 574.

40. In reply, Applicant amended the claims to recite the allowable feature of claim 6 (and intervening claim 5) in independent form as new independent claim 27 (and similar new independent claim 29 based on allowable claim 16). Applicant also amended independent claim 1 to partially include features of allowable claim 6 (and intervening claim 5) and amended independent claim 11 in a similar manner. *Id.*, 150-164. In the reply, Applicant distinguished the amended independent claim 1 (and 11) over the cited art, arguing that Kennewick⁵¹⁶ and Gong do not teach or suggest the feature "identifying a manner in which a natural language utterance was

spoken based on short-term knowledge and long-term knowledge about how a user utters a request.” *Id.*, 161-163.

41. The Examiner apparently rejected this argument, because the Notice of Allowance that subsequently issued included an Examiner’s Amendment authorized by Applicant’s representative in which independent claims 27 and 29 were cancelled, and independent claims 1 and 11 were further amended to recite additional features of allowable claim 6 (and intervening claim 5). *Id.*, 114-132.

42. In the Notice of Allowance, the Examiner made a Statement of Reasons for Allowance characterizing several references cited during prosecution as well as references cited during prosecution of related applications. *Id.*, 125-130, 248. The Applicant filed Comments responding to the Examiner’s statement submitting “that no one element or limitation in particular should be deemed to impart to, or be required for, patentability of the claims.” *Id.*, 4.

VII. RELATED LITIGATION AND RELATED PATENTS

43. I understand the ’699 patent is involved in IPR2025-00870.

44. I understand the ’699 patent is the subject of a pending litigation involving Petitioner and Patent Owner: *VB Assets, LLC v. Amazon.com Services, LLC*, No. 1:24-cv-00839 (D. Del.) (the “parallel litigation”). I understand Patent Owner sued Petitioner for alleged infringement of the ’699 patent and four other patents. *See* EX1013.

45. I understand the '699 patent has been involved or is currently involved in other litigations, including: *VB Assets, LLC v. SoundHound AI Inc.*, No. 1:24-cv-01279 (D. Del.); and *VB Assets, LLC v. Samsung Electronics Co. Ltd.*, No. 2:24-cv-00828 (E.D. Tex.).

46. I understand U.S. Patent No. 10,297,249, a parent of the challenged patent that shares the same specification, has been involved or is currently involved in litigations, including: the parallel litigation; *VB Assets, LLC v. SoundHound AI Inc.*, No. 1:24-cv-01279 (D. Del.); and *VB Assets, LLC v. Apple Inc.*, No. 1:24-cv-01368 (D. Del.).

47. I understand U.S. Patent No. 9,015,049, a parent of the challenged patent that shares the same specification, was involved in a prior litigation: *VB Assets, LLC v. Amazon.com, Inc.*, No. 1:19-cv-01410 (D. Del.). This patent was also involved in IPR2020-01346.

48. I understand that U.S. Patent No. 8,515,765, a parent of the challenged patent that shares the same specification, has been involved or is currently involved in other litigations including: *VB Assets, LLC v. Apple Inc.*, No. 1:24-cv-01368 (D. Del.); and *VB Assets LLC v. Samsung Electronics Co. Ltd.*, No. 2:24-cv-00828 (E.D. Tex.). This patent is also involved in IPR2025-00868.

49. I understand that U.S. Patent No. 8,073,681, a parent of the challenged patent that shares the same specification, was involved in multiple litigations and in the following IPRs: IPR2020-01367 and IPR2025-00866.

VIII. PRIOR ART TO THE CHALLENGED PATENT

50. I provide an overview of the state of the art at the time of the effective filing date and discuss below the specific references underlying the ground of invalidity set forth in this Petition.

A. Technological Background at Time of Filing

51. The '699 Patent relates to (i) a human user interacting with a system via speech recognition, and (ii) the system engaging with the user in a humanized conversational dialog . EX1001, 2:25-51. These ideas were well-known in the prior art at the time of the '699 Patent.

52. In interactive conversational systems involving a human user and a system, a dialog (or conversation) can be represented as a series of utterances, where each utterance is alternately generated by the human and by the system. A simple example of a dialog might be the following:

User: "I would like to fly to New York."

System: "When would you like to travel?"

User: "This Friday please."

System: "What is your departure airport?"

And so on. In this example there are two utterances by the user and two by the system. Technologies and systems to broadly support such interactive conversations had been under development and were well-known prior to 2006, both in academic research laboratories and in industry. *See, e.g.,* EX1009; EX1010; EX1011, 879-982 (Chapters 17-18).

53. These types of interactive conversational systems, as developed and implemented by 2006, relied on a number of related technologies that I will briefly describe below. The primary technologies are automated speech recognition (“ASR”), natural language understanding (“NLU”), dialog management (“DM”), and user modeling (“UM”). Broadly speaking, ASR involves recognizing words from spoken audio for each user utterance, NLU is the process of determining the meaning of the recognized words in the utterance, DM involves generation of appropriate response utterances by the system and the management of the conversation with the user across multiple utterances, and UM is the process by which the system keeps track of information about a user, e.g., what their typical goals and preferences are.

54. These terms are not the only ones used to describe these technologies. Other terms were (and are still) used by researchers and practitioners in this general context. One such term is “Spoken Language Understanding” (“SLU”) which typically encompasses aspects of ASR, NLU, and DM. Nonetheless, the terms ASR,

NLU, DM, and UM were in broad use in 2006 and would have been well-known to a POSITA at that time. Thus, I focus on these terms in my discussion since these terms are generally useful for describing the technical background for the challenged patent.

55. In 2006, ASR involved a sequence of steps consisting of (i) recording and digitizing speech waveforms spoken by a user, (ii) converting the digitized speech waveforms into numerical features that (typically) represent the frequency content of very short segments of the speech waveform, (iii) automatically classifying these features into corresponding phonemes (basic units of sounds within words), and (iv) recognizing specific words based on sequences of phonemes. This process has been the subject of extensive research and commercial activity since the 1960s. For example, the well-known text “SLP” devotes Chapters 5 through 13 (approximately 470 pages) just to ASR. EX1011, 227-711.

56. The output of the ASR component of a conversational interface system consists of the ASR’s best estimates of what words were spoken by a user in an utterance. The recognized words alone are typically not sufficient for an algorithm to understand the user’s goals and intent. Human language is very flexible and there are many ways to express the same meaning or intent. The technology of NLU consists of a broad range of algorithms and methods that allow a computer system to interpret the meaning of the recognized words from the ASR. EX1011, 893-912.

For example, consider the simple phrase “I would like to fly to New York” from the dialog example described earlier. And assume that the system is a travel reservation system that can find flights, make flight reservations, pay for the flights, suggest or advertise additional travel services, and so on. Such a system would have an internal representation (e.g., a database) of items such as names of airports, locations of airports, lists of flights, actions the system can take (e.g., book flights), and so on. Thus, it would need to interpret the user’s words “like to fly” as an implicit request for the system to find a flight. It would also need to interpret the word “I” to mean that the flight is for the individual who is speaking. And it would need to understand that “New York” is the name of a location, and that it is not the name of a specific airport but is a city with multiple airports associated with it. There are different approaches associated with developing NLU technologies that can robustly handle such situations, and by 2006 multiple such systems had been developed that demonstrated the necessary NLU capabilities for interpreting the meaning of spoken words in conversational dialog. *See, e.g.,* EX1011, 945-982 (Chapter 18); EX1010; EX1012, 18:39–19:37.

57. As I describe above, the ASR and NLU components of a conversational interface system focus on recognizing words and interpreting these words on a per-utterance basis. DM (Dialog Management) technology addresses the problem of tracking different aspects of a dialog across multiple utterances. For example, for a

travel reservation system interacting with a user via speech, the DM component of the system gradually accumulates the relevant pieces of information from the user (dates, times, arrival/departure airports, seating preferences, confirmation of booking, purchase information) over the course of a dialog, i.e., over a sequence of utterances generated by the user. Thus, DM technologies involve creating internal representations of concepts such as goals for a user and methods to satisfy those goals. One such method includes directly asking the user for specific information, which in turn involves automatically generating the appropriate words for the system's response (utterance) to the user, given what the system knows already (e.g., "What time would you like to leave for your flight to New York?") and then converting that into a spoken response that the user can hear. Another type of DM method involves inferring the relevant information from knowledge sources available to the system. For example, the DM might assume from information in the user's profile that the departure airport is likely to be the one closest to where the user lives. And the DM could infer the likely destination airport if non-stop flights from the user's home airport only go to a specific airport in the New York area. DM technology involves various software components that can make decisions about what type of information to request from the user, how to present this information to the user, in what sequence to request this information, when this information might be inferred from other sources, and so on. DM technology that exhibited these

characteristics was well-known by 2006 and had been developed in research laboratories and in commercial systems before that time. *See, e.g.*, EX1011, 912-927 (Chapter 17.6-17.7), 945-982 (Chapter 18); EX1010; EX1004.

58. It was also well-known by 2006 that there were significant benefits to be gained if a conversational system had access to specific knowledge about a user (such as their travel preferences). Such access can make the interaction with the user more efficient from the system's perspective and more pleasant from the user's point of view, by reducing the amount of information the system needs to request from the user during the dialog. The technology of user modeling (UM) consists of different methods the system uses to automatically build an internal representation of a user. These methods can include simple (but useful) approaches such as recording information about a user's full name, age, home address, etc., explicitly provided by a user during (say) a registration process. More complex UM methods extract information that is implicitly provided by the user, for example, inferences about a user's travel preferences from words in past conversations that a user has had with the system, or inferences from words in earlier utterances in the current conversation. By 2006, the benefits of UM in conversational modeling were widely acknowledged and had been developed in the context of both research and commercial systems. *See, e.g.*, EX1004.

59. Humanizing interactions between a conversational system and user was also a well-known concept by 2006. Researchers understood that making a dialogue system feel more natural required the system to track the evolving subject matter of a conversation and to deprioritize or discard stale information once the conversation had moved on. For example, Matsuda describes a conversation system that “makes it possible to achieve a natural conversation, such as with a conversation between people.” EX1007, ¶116. Matsuda’s system adapts to changes in what the user is talking about and dynamically shifts focus away from subjects that have not been referenced for a period of time, reducing the likelihood of generating responses that are no longer relevant. This approach mirrors how people naturally conduct conversations, by focusing on the current topic rather than using outdated information when responding to an utterance. EX1007, ¶¶84, 100, 116.

60. Thus, all the foundational technologies referenced in the claims were already known to a POSITA at the time of the effective filing date of the ’699 Patent.

B. Summary of the Prior Art

1. Kennewick (EX1004)

61. Kennewick, U.S. Patent App. Pub. No. 2004/0193420, is entitled “mobile systems and methods for responding to natural language speech utterance.” Kennewick was filed July 15, 2003 and published September 30, 2004. EX1004. I understand Kennewick is prior art at least because its publication date predates the

October 16, 2006 effective filing date of the challenged patent by more than one year. I form no opinion as to whether Kennewick is entitled to an earlier priority date.

62. Kennewick describes a system that allows “users to ask natural language speech questions” and “present[s] results in a natural manner.” EX1004, ¶2. The described system is “a complete speech-based information query, retrieval, presentation and command environment” that “makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands.” *Id.*, ¶11. In particular, Kennewick describes “a mobile interactive natural language speech system...that includes a speech unit.” *Id.*, ¶12. “[A] computer device or system...receives input from the speech unit and processes the input (e.g., retrieves information responsive to the query, takes action consistent with the command and performs other functions as detailed herein), and responds to the user with a natural language speech response.” *Id.*, ¶14.

63. Kennewick explains that a “dialog with the system may begin when a user first addresses it” and that “[a]t the end of a dialog or to interrupt a dialog, the user may speak a dismissal word (‘good by’).” *Id.*, ¶156. Kennewick discloses tagging words and phrases recognized for each utterance of the dialog, where “the tags, added by the speech recognition engine 120 to each utterance, may allow other

components of the system to tie that utterance to the correct . . . dialog”. *Id.*, ¶155. Referring to the tagged information as both “user dialog history” and “history of the dialog” interchangeably, Kennewick discloses using the “dialog history” or the “history of the dialog” to generate an appropriate response. *Id.*, ¶¶157, 160, 162, 164, 170.

64. Kennewick further discloses that its “invention creates, stores and uses extensive personal profile information for each user” to “improv[e] the reliability of determining the context and presenting the expected results for a particular question or command.” *Id.*, Abstract, ¶31; *see also id.*, ¶160 (describing “determin[ing] a context for an utterance” based on a “user profile,” among other values). Kennewick discloses that its “user profile” includes, among other things, “history of questions asked,” “session histories,” “history of commands and queries,” and “history of [the user’s] interactions with the system.” *Id.*, ¶¶31, 143. Information in the user profiles “may be added and updated automatically as the user uses the system,” or “learned by the system based on user behavior.” *Id.* Kennewick discloses using the “user profile” to generate an appropriate response. *Id.*, ¶¶157, 160, 162, 164, 170. Kennewick’s system further formats the response to “simulate some aspects of human ‘personality,’” so that responses “do not appear rigidly formatted or mechanical,” but instead can exhibit traits such as “sarcasm, humor, irritation, and sympathy” *Id.*, ¶¶24, 141.

2. Cooper (EX1005)

65. Cooper, U.S. Patent No. 6,757,362, is entitled “personal virtual assistant.” Cooper issued June 29, 2004. EX1005. I understand Cooper is prior art at least because its issue date predates the October 16, 2006 effective filing date of the challenged patent. I form no opinion as to whether Cooper is entitled to an earlier priority date.

66. Cooper describes a “virtual assistant [that] adapts to the user in many different ways based on the input the virtual assistant receives.” EX1005, 2:56-64, Abstract. Cooper determines the user’s manner of speech, e.g., by identifying when “words associated with polite discourse are included in input received from the user.” *Id.*, 43:55-59. Words associated with polite discourse could include, “for example, ‘please,’ thank you,’ ‘thanks,’ ‘excuse,’ ‘pardon,’ ‘may I,’ ‘would you mind,’ or other words and phrases associated with polite discourse.” *Id.*, 43:51-55. “If, such words of polite discourse are included in the input received from the user, the virtual assistant could automatically enable the politeness setting.” *Id.*, 43:59-61. A “politeness setting . . . when enabled, would cause the virtual assistant to include words or phrases associated with polite discourse in the output from the virtual assistant.” *Id.*, 43:48-51. Thus, Cooper’s automatically-enabled politeness setting ensures that the responses from the virtual assistant will match the manner of speech of the user’s utterances.

67. As another example, Cooper describes “the user information input into the virtual assistant could be information about the user’s emotion, which could be based on information about the user’s voice volume, word choice and speech rate. Based on such information, the virtual assistant could automatically determine the user’s emotional state, calm or angry, for example. If the user’s emotional state is angry, the output of the virtual assistant could automatically include words associated with submissive discourse, such as ‘sorry,’ ‘regret’ and ‘apologize.’” *Id.*, 43:62-44:4. Moreover, the “virtual assistant could save the information obtained about the user’s emotional state for use in future sessions.” *Id.*, 44:4-6. Thus, Cooper’s automatic inclusion of submissive discourse upon detection of an angry emotional state in the user input ensures that the responses from the virtual assistant will be appropriate given the manner of speech of the user’s utterances.

3. Matsuda (EX1006; EX1007)

68. Matsuda is Japanese Laid Open Patent Publication No. JP2004334591. Matsuda was filed May 8, 2003 and published November 25, 2004. EX1006. Exhibit 1007 is a certified English translation of Matsuda. I understand Matsuda is prior art at least because its publication date predates the October 16, 2006 effective filing date of the challenged patent. I form no opinion as to whether Matsuda is entitled to an earlier priority date.

69. Matsuda is titled “Conversation System and Conversation Processing Program.” EX1007, 1. Matsuda describes a conversation system that uses a “subject person number” to track a subject (e.g., a person) currently being discussed in a conversation. *Id.*, ¶26. This subject person number is reset if “at least a prescribed amount of time (for example, 10 minutes) has elapsed without a name of that person being mentioned.” *Id.*, ¶¶26, 38. In that case, “the CPU 11 sets the ‘subject person number’ to 0 [] and clears the ‘subject person number latest setting date/time.’” *Id.*, ¶38. Matsuda explains that this enables the system to achieve a more natural conversation with the user, by being able recognize subject names referred to by the user earlier in the conversation, even when the user does not use the “exact same name.” *Id.*, ¶¶84, 97.

70. Matsuda also discloses that name-related information is retained across conversations in a persistent data structure (“name data 20a”). *Id.*, ¶31. When the user utters a name (i.e., subject or topic), the system compares the spoken name against existing entries in the data structure for that user and if a match is found, sets the person number of an individual identified in a matching entry as the current subject. *Id.*, ¶76. If no match is found, the system determines that the person has not been previously registered. *Id.*, ¶74. The system uses different sets of rules when determining whether the user’s spoken name corresponds to a previously registered name, depending on whether there is a current “subject person” for the

conversation. *See id.*, ¶¶77-82. Matsuda further updates the data structure based on the user's utterances, so that "even in conversations on following days," the user may use "the names that he or she always uses" to refer to particular subjects. *Id.*, ¶¶7, 78-79, 101.

71. Accordingly, Matsuda discloses using the "subject person number" and associated timestamp indicating an amount of elapsed time since the subject was last mentioned by the user to humanize a conversation by enabling the system to maintain focus on a current subject while also recognizing when "the conversation [has] moved away from the subject of that person." *Id.*, ¶¶26, 38, 100. This approach mirrors how people naturally conduct conversations, by focusing on the current topic rather than using outdated information when responding to an utterance. EX1007, ¶¶84, 100, 116.

IX. INVALIDITY

72. In my opinion, Kennewick in combination with Cooper and Matsuda renders obvious claims 1-22 of the '699 patent. My opinion regarding invalidity of the '699 patent is based on the understanding of a POSITA in the relevant timeframe and my analysis of the references. I evaluated the limitations of each claim as a whole, and did not merely look for where each individual limitation could be found in the prior art. I also considered what a POSITA might understand from the prior

art reference and what modifications and alterations a POSITA would know could be made to the prior art reference.

73. I explain my opinions in more detail below. I have included the language of each claim below for reference purposes, but based my opinions on the claims as they appear in the patent; any differences between the excerpted claim language below and the patent are unintentional.

A. Claims 1-11 are Rendered Obvious by Kennewick, Cooper, and Matsuda

1. Claim 1

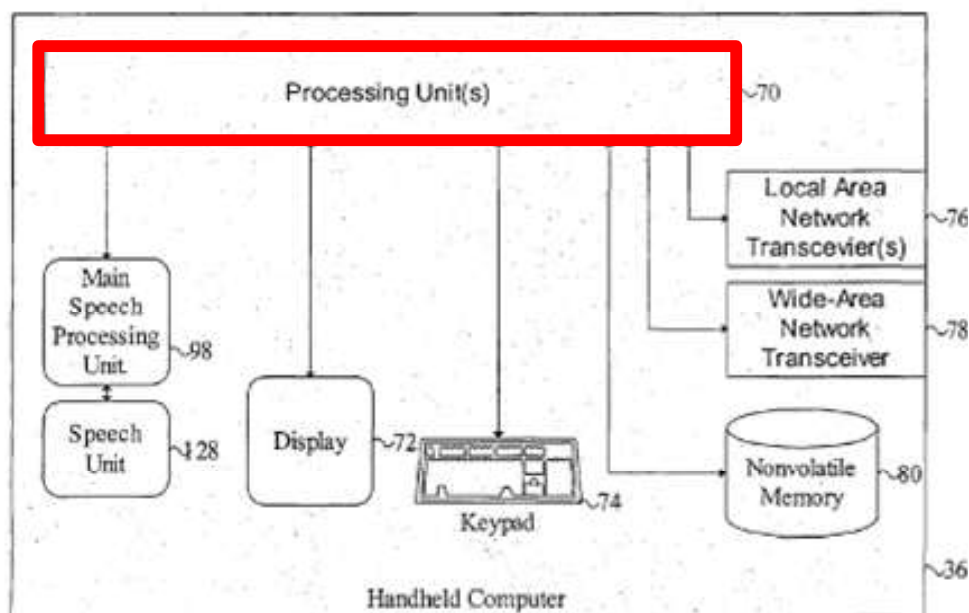
[1.pre] “A computer-implemented method of generating natural language system responses adapted based on a user's manner of speaking, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:”

74. Kennewick in combination with Cooper discloses the preamble. Kennewick relates to an “integrated environment allowing mobile users to ask natural language speech questions or give natural language commands in a wide range of domains.” EX1004, ¶2. The environment “present[s] results in a natural manner even in cases where the question asked or the responses received are incomplete, ambiguous or subjective.” *Id.*, ¶¶2, 23-24 (explaining “the system may simulate some aspects of human ‘personality’”), 101, 155-56 (disclosing a “dialog with the system”).

75. A POSITA would have understood that Kennewick’s environment performs a “*method of generating natural language system responses.*” Kennewick discloses that the environment is provided by “systems and methods.” *Id.*, ¶11. A POSITA would have understood Kennewick to disclose that the systems it describes perform the described methods, including a “*method of generating natural language system responses*” that receives natural language utterances and responds in kind. *See, e.g., id.*, ¶¶14 (the system “responds to the user with a natural language speech response”), 27 (“[t]he system, according to one aspect of the invention, can process and respond to questions, queries and commands”), 28 (“[t]he invention can be used for generalized local or network information query, retrieval and presentation in a mobile environment. For each user utterance including a question or query or set of questions or queries, the system may perform multiple steps”), 118 (“[t]he natural language interactive speech processing system may make maximum use of context, prior information, location information, domain information and user specific profile data to achieve a natural environment for one or more users making queries or stating commands across multiple domains”), claim 28 (“[a] method responsive to a user generated natural language speech utterance”).

76. Kennewick also discloses that the method is “*computer-implemented.*” Kennewick’s disclosed method may be implemented on numerous “computer alternatives” that Kennewick collectively refers to as “a computer.” *Id.*, ¶13. The

computer may be a “fixed computer” or “handheld device.”² *Id.* Kennewick’s Figure 3 illustrates a possible embodiment of the handheld computer.



Id., Figure 3, ¶90 (annotated).³ Kennewick’s Figure 4 is a near-identical figure illustrating an embodiment of fixed computer. *Id.*, Figure 4, ¶91.

² Kennewick uses the terms “handheld,” “handheld device,” and “handheld computer” interchangeably, often with the same reference numeral 36. *See, e.g.*, EX1004, ¶22 (using terms “handheld device,” “handheld computer,” and “handheld”), 101 (“handheld devices 36”), 102 (“handheld computers 36”), 110 (“a handheld computer 36 can be used as a component of the invention”).

³ All annotations and emphases are added unless otherwise noted.

77. A POSITA would have understood that these embodiments of the computer are examples of “***a computer system***” because they are formed of various computing components working together. *See id.*, ¶3 (“[t]elematics systems are systems that bring human-computer interfaces to vehicular environments”), 96 (“[a] speech unit 128 can be permanently attached to the vehicle 10 or can be part of a handheld device 36 or a fixed home or office computer system 44”), 107 (“the main speech processing unit 98 and speech unit 128 can be embedded in one or more of the TCUs 28, handheld computers 36 and fixed computer systems 44”). Therefore, a POSITA would have understood that Kennewick’s “***computer-implemented method***” is “***implemented by a computer system.***”

78. Kennewick’s computer system “***includes one or more physical processors.***” Figure 3, shown above, illustrates that the computer has “**processing units.**” These “may include one or more central processing units, one or more data and address busses, data interfaces and volatile memory.” *Id.*, ¶¶111, 115 (similar), 101, 122, 107-09. A POSITA would have understood that the “central processing units” are “***one or more physical processors***” because physical processors acting as central processing units are conventional.

79. A POSITA likewise would have understood that the physical processors “***execut[e] one or more computer program instructions which, when executed, perform***” Kennewick’s method. Kennewick discloses that “[o]ne aspect

of the invention includes software that is installed onto the computer,” where the software includes a “a speech recognition module for capturing the user input,” “a parser for parsing the input,” and “a text to speech engine module for converting text to speech.” *Id.*, ¶13. Kennewick also discloses that “domain specific behavior and information are organized into agents” which “are executables that receive, process and respond to user questions, queries and commands.” *Id.*, ¶17. A POSITA would have understood that Kennewick’s modules and agents are computer program instructions which Kennewick’s physical processors execute to perform the method because this is how computers conventionally operate. *See, e.g., id.*, ¶17 (“[a]gents can be packages of executable code”), 107 (“the main speech processing unit 98 and speech unit 128 can be embedded in...the...handheld computers 36 and fixed computer systems 44”).

80. For these reasons, a POSITA would have understood Kennewick to disclose “[a] *computer-implemented method of generating natural language system 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.*”

81. Kennewick also discloses using different personalities for responses. Kennewick explains that “[e]xamples of personality traits include, sarcasm, humor, irritation, and sympathy, and other traits.” *Id.*, ¶141. Kennewick’s system may

simulate different response personalities, stating that the “use of simulated personality characteristics is also desirable” and giving examples such as presenting upsetting information “in a sympathetic manner” or urgent alerts “with a definite and authoritative personality.” *Id.*, ¶24. The system may dynamically select which personality characteristics to use for responses based on the “context” of the conversation as well as user profile information. *Id.*, ¶179.

82. Further, Kennewick discloses that its agent-based dialog system tailors responses based on both domain knowledge and user preferences. EX1004, ¶¶17, 23, 24. Kennewick states that “domain specific behavior and information are organized into agents” and that “[a]gents are executables that receive, process and respond to user questions, queries and commands.” *Id.*, ¶17. These agents “include everything that is needed to extend the functionality of the invention to a new domain.” *Id.* Further, Kennewick teaches that response presentation is influenced by user preferences: “The optimal formatting and presentation depends on the context of the queries, the contents of the response being presented, the history of the interaction with the user, the user's preferences and interests and the nature of the domain.” *Id.*, ¶23.

83. Kennewick does not explicitly disclose that the generated responses are “*adapted based on a user’s manner of speaking.*” This aspect is taught by Cooper.

84. Cooper describes a “virtual assistant [that] adapts to the user in many different ways based on the input the virtual assistant receives.” EX1005, 2:56-64, Abstract. As Cooper explains, its virtual assistant can consider “whether words associated with polite discourse are included in input received from the user.” *Id.*, 43:55-59; 2:26-30 (criticizing prior art’s failure to automatically adapt when “user’s input contains words associated with polite discourse”). Words associated with polite discourse could include, “for example, ‘please,’ thank you,’ ‘thanks,’ ‘excuse,’ ‘pardon,’ ‘may I,’ ‘would you mind,’ or other words and phrases associated with polite discourse.” *Id.*, 43:51-55. If “words of polite discourse are included in the input received from the user, the virtual assistant could automatically enable the politeness setting.” *Id.*, 43:59-61. Enabling the politeness setting “would cause the virtual assistant to include words or phrases associated with polite discourse in the output from the virtual assistant.” *Id.*, 43:49-51.

85. Cooper also explains that “the user information input into the virtual assistant could be information about the user’s emotion, which could be based on information about the user’s voice volume, word choice and speech rate. Based on such information, the virtual assistant could automatically determine the user’s emotional state, calm or angry, for example.” *Id.*, 43:62-67; *see also id.*, 44:1-4 (“[i]f the user’s emotional state is angry, the output of the virtual assistant could

automatically include words associated with submissive discourse, such as ‘sorry,’ ‘regret’ and ‘apologize.’”).

86. A POSITA would have understood that Cooper discloses generating responses “*adapted based on a user’s manner of speaking*” because Cooper’s virtual assistant “changes” and “adapts” its output based on how the user speaks to it. *See* EX1005, Abstract. This understanding is consistent with disclosure in the ’699 patent, which describes “adapt[ing] 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.” EX1001, 6:32-34; *see also id.*, 18:45-50 (the “[a]daptive 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).”).

87. A POSITA would have found it obvious to combine Cooper with Kennewick such that Kennewick’s computer system generates natural language system responses “*adapted based on a user’s manner of speaking*.” Kennewick and Cooper are analogous art; both relate to voice-based interactions with computers. *See, e.g.,* EX1004, ¶2 (“[t]he present invention relates to the retrieval of online information and processing of commands through a speech interface”);

EX1005, 2:37-39 (“one aspect of the present invention is a computer-based virtual assistant...comprising a voice user interface”).

88. Kennewick’s dialog system uses different personalities for responses, giving examples such as presenting upsetting information “in a sympathetic manner” or urgent alerts “with a definite and authoritative personality.” EX1004, ¶24. Kennewick’s responses are further tailored based on domain knowledge and user preferences. *Id.*, ¶¶17, 23, 24. Likewise, Cooper teaches that its virtual assistant automatically enables the politeness setting if words of polite discourse are included in the input received from the user. EX1005, 43:59-61. Enabling the politeness setting “would cause the virtual assistant to include words or phrases associated with polite discourse in the output from the virtual assistant.” *Id.*, 43:49-51; *see also id.*, 44:1-4 (“[i]f the user's emotional state is angry, the output of the virtual assistant could automatically include words associated with submissive discourse, such as ‘sorry,’ ‘regret’ and ‘apologize.’”). Cooper’s “virtual assistant adapts to the user in many different ways based on the input the virtual assistant receives,” and Cooper lists such input to include “whether the user uses words associated with polite discourse” and “the user's emotional state.” *Id.*, 2:56-63.

89. A POSITA would have found it obvious to incorporate Cooper’s teachings regarding recognizing and responding appropriately to the user’s manner of speech including, for example, (1) automatically enabling a politeness setting

when it recognizes words associated with or containing polite discourse in the speech input received from the user; and/or (2) automatically determining a user's emotional state (e.g., calm or angry) in the speech input received from the user and automatically adapting the output based on the determined emotional state, into Kennewick's dialog system that uses different personalities for responses and tailors responses based on domain knowledge and user preferences. A POSITA would have been motivated to make this combination because adapting Kennewick's dialog system to tailor its responses not just based on domain knowledge and user preferences but also based on the user's manner of speech (e.g., emotional state or polite discourse cues) as taught by Cooper would improve the naturalness and appropriateness of system responses. This is consistent with Kennewick's stated objective of making "responses to the user's questions and commands seem more natural." EX1004, ¶¶24, 35.

90. This is a combination of prior art elements according to known methods to yield the predictable result of adapting the response from the computer system to the user's manner of speaking thereby making it more natural and human-like. The combination is also a predictable use of known techniques for improving voice-based interactions with computers. The improvement would have been within the skill of a POSITA, who would reasonably expect success based on the compatibility of the systems and their shared goal of enabling adaptive, conversational interfaces.

[1.a]: “receiving, by the computer system, a user input comprising a natural language utterance;”

91. Kennewick discloses limitation [1.a]. Kennewick’s “mobile interactive natural language speech system,” which is the “**computer system**,” includes a **speech unit** (speech interface device). EX1004, ¶12 (“a mobile interactive natural language speech system (herein ‘the system’) is provided that includes a speech unit”).

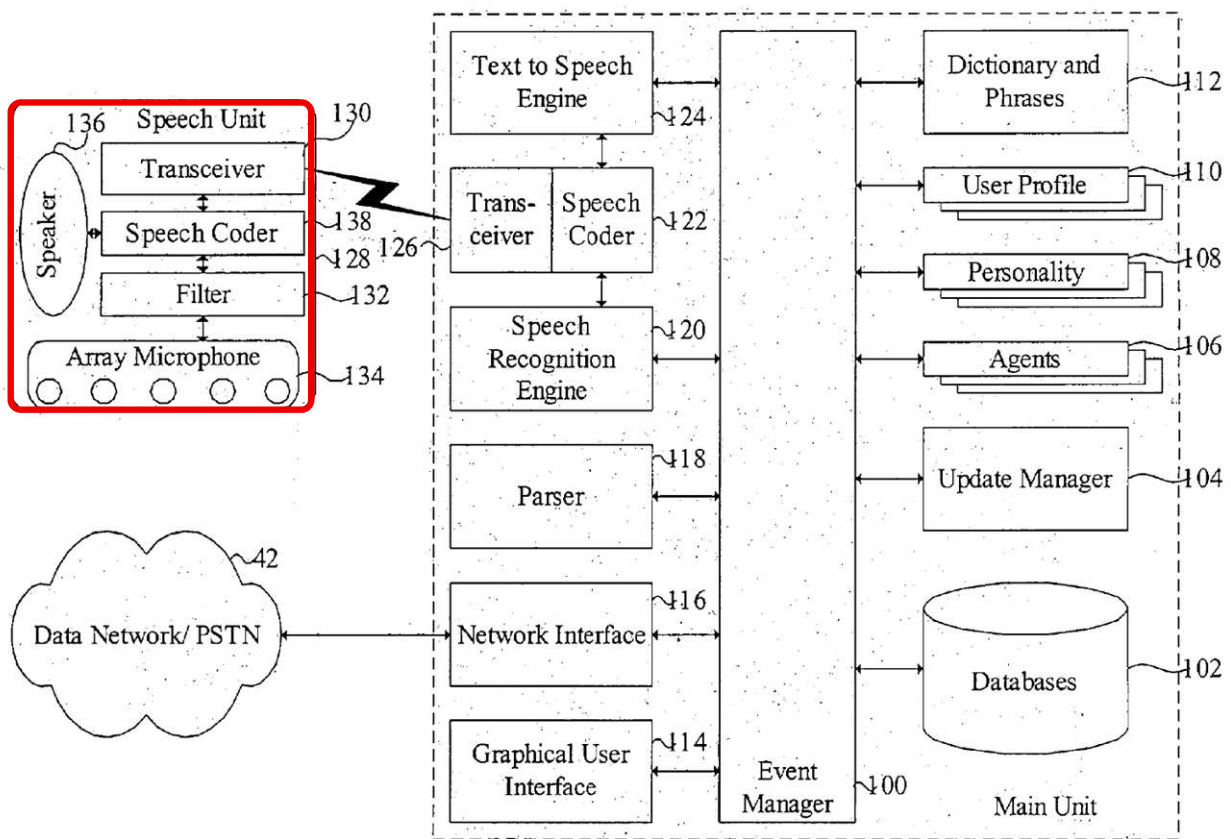


Figure 5. Speech Processing System Block Diagram

Id., Fig. 5.

92. The **speech unit** receives “a user input comprising a natural language utterance.” Kennewick explains that the “**speech unit 128** includes one or more microphones, for example array microphone 134, to receive the utterances of the

user.” *Id.*, ¶121. These utterances are *a user input comprising a natural language utterance.*” *See id.*, ¶14 (“the system may include a speech interface device that receives spoken natural language queries, commands and/or other utterances from a user”). *See also id.*, ¶¶2 (“the invention is a fully integrated environment allowing mobile users to ask natural language speech questions or give natural language commands in a wide range of domains”), 34 (“the invention is intended to be able to accept most any natural language question or command”), 73 (“[v]ehicle operators and other occupants can use the interactive natural language interactive speech interface of the invention to perform many types of information query, retrieval and presentation operations”), 101 (“[i]n every case, the invention may provide an interactive natural language speech user interface to the services offered”), 152, claims 1, 3, 28, 43 (reciting variations of “receiving the user generated natural language speech utterance”).

[1.b] “recognizing, by the computer system, one or more words or phrases from the natural language utterance;”

93. Kennewick discloses limitation [1.b]. The natural language utterances received by the **speech unit** are coded by the speech coder 138 within the speech coder and transmitted to the main unit 98. *Id.*, ¶121. From there, the coded speech is “passed to the speech coder 122 [of the main unit 98] for decoding and decompression.” *Id.*, ¶123.

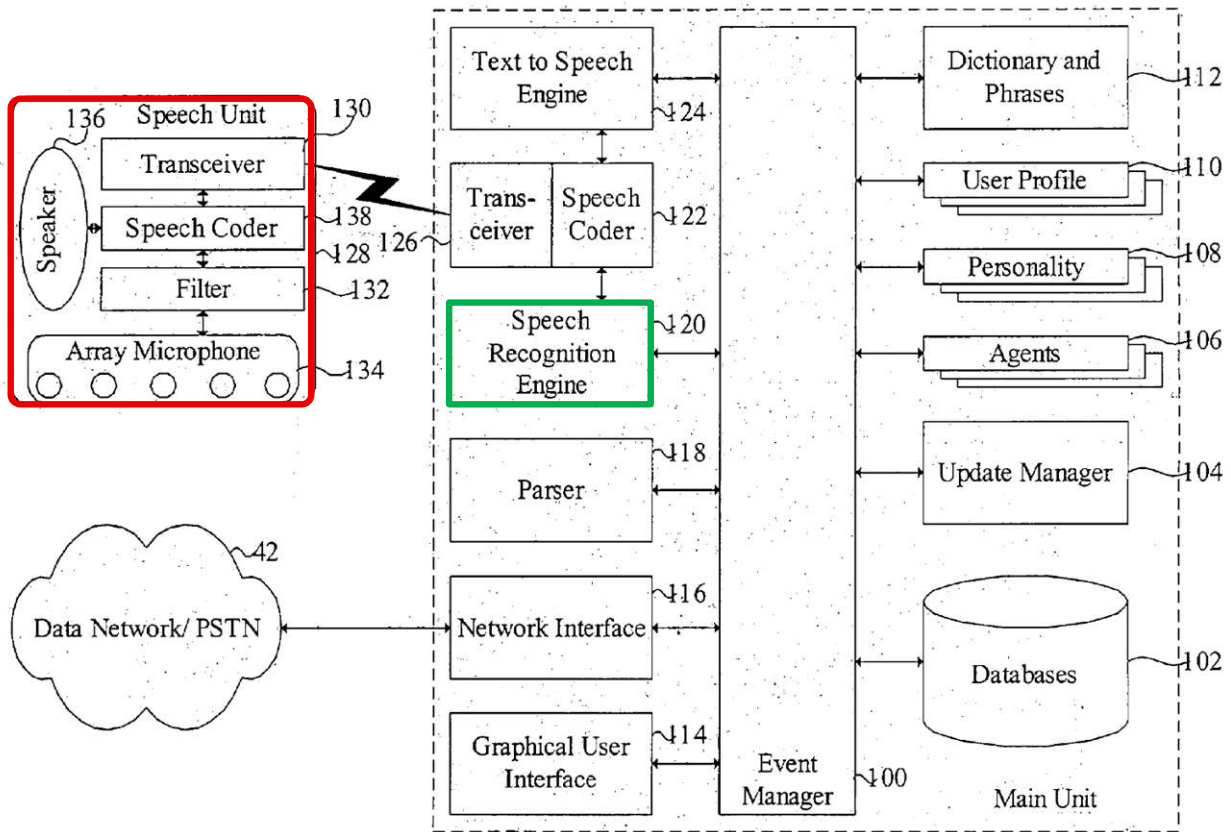


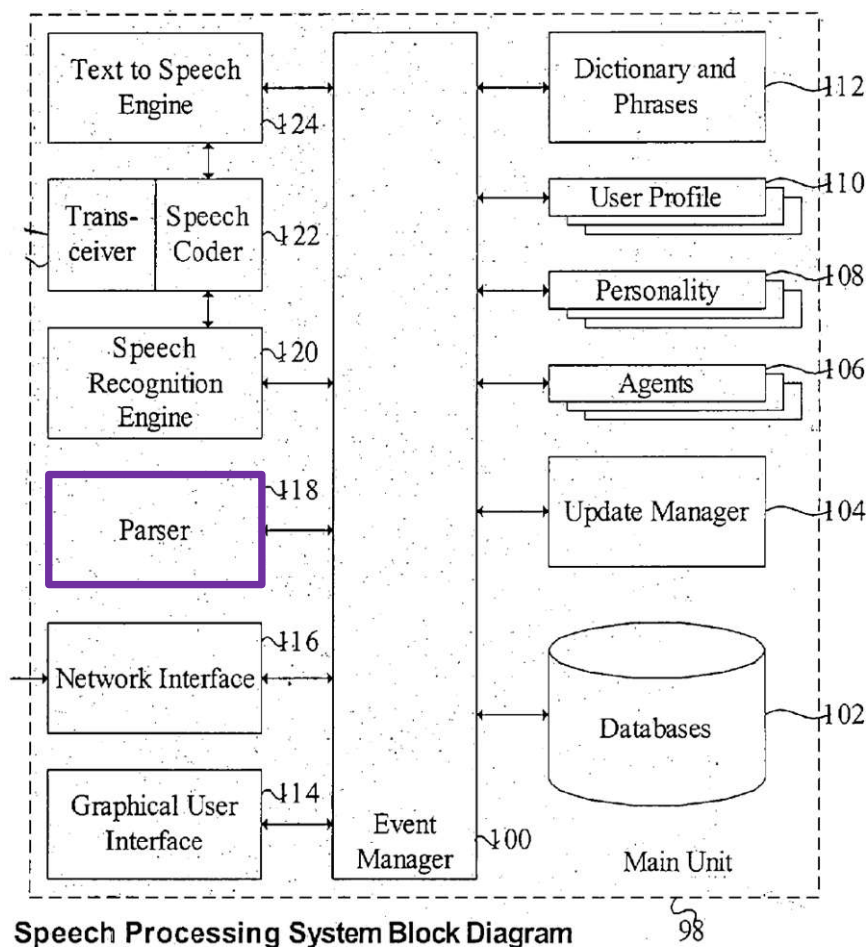
Figure 5. Speech Processing System Block Diagram

94. The “output of the speech coder 122 may be fed to the **speech recognition engine 120**. EX1004, ¶155, Fig. 5. The speech recognition engine 120 may recognize words and phrases, using for example, information in the dictionary and phrase tables 112, and pass these to the parser 118 for interpretation.” *Id.*, ¶155; *see also id.*, ¶123 (“decoded speech can be processed by the speech recognition engine 120 using data in the dictionary and phrases module 112 and received from the agents 106. The recognized words and phrases may be processed by the parser 118, which transforms them into complete commands and questions using data supplied by the agents.”).

95. A POSITA would have understood that Kennewick's speech recognition engine 120 recognizing the "words and phrases" in the received "utterances of the user" meets the claimed "*recognizing, by the computer system, one or more words or phrases from the natural language utterance.*" *Id.*, ¶¶121, 123, 155.

[1.c] "identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;"

96. Kennewick discloses limitation [1.c]. Kennewick discloses a "*computer system*" receiving a "*natural language utterance*" as I discuss for limitation [1.a]. As I discuss for limitation [1.b], Kennewick's computer system also recognizes "*one or more words or phrases...from the natural language utterance.*" EX1004, ¶¶123, 155; *see also id.*, Figure 5. Kennewick further teaches that its computer system identifies "*a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance.*"



EX1004, Figure 5 (excerpt).

97. Kennewick explicitly discloses that its **parser 118** determines (identifies) a context based on the words and phrases recognized by the speech recognition engine 120. EX1004, ¶160. According to Kennewick: “[o]nce the words and phrases have been recognized by the speech recognition engine 120, the tokens and user identification may be passed to the parser 118. The parser examines the tokens for the questions or commands, context and criteria.” *Id.* Kennewick further explains that “[t]he parser may determine a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching.” *Id.* Thus,

the parser 118 in Kennewick identifies the “context” “based on” the recognized “tokens” – i.e., recognized words or phrases from the user’s natural language utterance. *Id.*

98. Accordingly, a POSITA would have understood Kennewick to disclose and render obvious “*identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance.*”

[1.d] “determining, by the computer system, an interpretation of the natural language utterance based on the identified context;”

99. Kennewick discloses limitation [1.d]. Kennewick discloses a “*computer system*” receiving a “*natural language utterance*” as I discuss for limitation [1.a]. As I discuss for limitation [1.c], Kennewick’s computer system identifies “*a context*” for the natural language utterance. EX1004, ¶160.

100. Kennewick also discloses “*determining, by the computer system, an interpretation of the natural language utterance based on the identified context.*”

Id., ¶¶123, 155, 160, 162-63, 165. Kennewick’s system includes a parser that interprets natural language utterances received from the user. *Id.*, ¶¶28 (“[f]or each user utterance including a question or query or set of questions or queries, the system may perform multiple steps possibly including...parsing and interpreting the question or query”), 41 (“[f]or each user command utterance, the system may execute a number of steps possibly including...parse and interpret the command”),

143 (“[c]ommands and questions are interpreted”), 155 (“[t]he speech recognition engine 120 may recognize words and phrases...and pass these to the parser 118 for interpretation”).

101. The parser determines the interpretation based on the context of the utterance. *Id.*, ¶7 (“some questions can only be interpreted in the context of previous questions”). In particular, the parser uses recognized words and phrases, context, and other criteria to determine the appropriate context-based agent to be invoked for a user utterance forming a question or command. *Id.*, ¶¶160 (“[t]he context of a question or command may determine the domain and thereby, the domain agent 156, if any, to be invoked”), 162 (“[o]nce the context for the question or command has been determined, the parser 118 can invoke the correct agent”), 165 (“[b]ased on the context, the parser can evoke the correct agent to process the question or command”). The invoked agent works with the parser to interpret the utterance based on the context. For example, if the user asks “what about flight one hundred and twenty too?,” the parser may determine that the context is “travel planning” and invoke a travel planning domain agent specific to this context, where “[t]he parser and domain agent use flight information in the database and network information along with context to determine the most plausible interpretation among[] flight 100 and flight 20 also, flight 100 and flight 22, flight 122, etc.” *Id.* ¶163.

102. Thus, because Kennewick describes its computer system determining an interpretation of the utterance by invoking an agent based on context, a POSITA would have understood that Kennewick discloses and renders obvious *“determining, by the computer system, an interpretation of the natural language utterance based on the identified context.”*

[1.e] “accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;”

103. Kennewick in combination with Matsuda discloses and renders obvious limitation [1.e].

104. Kennewick discloses *“a single conversation between a user and the computer system.”* Kennewick discloses that the user and the computer system have a “dialog.” The dialog starts with the user addressing the computer system and ends with the user uttering a dismissal word. EX1004, ¶156 (a “dialog with the system may begin when a user first addresses it” such as “by speaking a generic word (‘computer’) and/or addressing a specific name (‘Fred’).”); *see also id.*, (“At the end of a dialog or to interrupt a dialog, the user may speak a dismissal word (‘good by’).”). A POSITA would have understood and found obvious from Kennewick’s disclosure that the dialog occurring between the addressal and dismissal words is a *“single”* dialog because the words indicate the start and end of one dialog. A

POSITA also would have understood that this dialog is a “**conversation**” between the user and the computer system. *See, e.g.*, EX1008 (dictionary defining “dialogue or dialog” as “A conversation between two or more people.”).

105. Kennewick further discloses that the computer system receives “**one or more natural language utterances...related to a single conversation between a user and the computer system.**” Kennewick discloses that the computer system receives a user input comprising a natural language utterance as I discuss for limitation [1.a]. A POSITA would have understood that the user input is related to a single conversation between the user and the computer system because the computer system receives the input during a conversation. For example, the computer system may receive within a single conversation a natural language utterance from the user indicating a question or command and additional user utterances resolving any ambiguities in the question or command. *See, e.g., id.*, ¶161 (“the system can ask a question of the user to verify the question or command is correctly understood”). Therefore, Kennewick discloses receiving “**one or more natural language utterances**” “**related to a single conversation between a user and the computer system.**”

106. Kennewick also discloses the “**computer system**” “**accumulating...short-term knowledge based on one or more natural language utterances...related to a single conversation between a user and the computer**

system.” The ‘699 patent describes “short-term knowledge” as knowledge that “may accumulate during a single conversation, where input received during a single conversation may be retained.” EX1001, 5:10-12, 14:5-6, 16:59-62 (“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”), 17:26-29 (“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)”).

107. Kennewick, in turn, describes that during a dialog (a conversation) with the user, the method adds “tags . . . to each utterance [in the dialog] . . . to tie that utterance to the correct . . . dialog” and user. EX1004, ¶¶155-56. To accomplish such tagging, the method first “recognize[s] words and phrases, using for example, information in the dictionary and phrase tables 112” and “determine[s] the user’s identity by voice and name for each utterance.” *Id.* The method then tags “[r]ecognized words and phrases” “with this identity in all further processing” of the dialog. *Id.* A POSITA would have understood that Kennewick’s system tagging recognized words and phrases for each utterance in order to tie that utterance to the correct dialog results in “*accumulating*” a dialog history over time.

108. A POSITA also would have understood that Kennewick’s dialog history is “*short-term knowledge*” because it reflects details about a user’s

utterances received during a conversation. Indeed, Kennewick’s description of building of a dialog history by systematically tagging each utterance of the dialog (e.g., *id.*, ¶155) is consistent with the ’699 patent’s own disclosure of accumulating “short-term knowledge” including a “history of the conversation,” which is accumulated “as the conversation progresses.” See EX1001, 16:50-53 (“[f]or 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”). Accordingly, a POSITA would have understood Kennewick to disclose ***“accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received ..., wherein the one or more natural language utterances received ... are related to a single conversation between a user and the computer system.”***

109. However, Kennewick does not explicitly disclose ***“accumulating...short-term knowledge based on one or more natural language utterances received during a predetermined time period.”*** This aspect is taught by Matsuda.

110. Matsuda discloses a “conversation system” such as a “voice-interaction robot” to “enable a user to ... convers[e] with a computer.” EX1007, ¶1. Matsuda’s system takes as input “voice from the user during the conversation,” and generates

“a response to a statement by the user.” *Id.*, ¶22. Matsuda’s system stores name data that tracks names of people who are the subjects of the user’s current or past conversations. *Id.*, ¶¶13 (“In a conversation system of this configuration, the specific subject is a person”), 26, 31 (“The name data...are data indicating names related to individual persons, stored for each individual conversation participant”), Figures 10-12. The name data for a person also includes an associated “person number” that identifies the person to the system, and may include one or more names used by the user to refer to the person. *Id.*, ¶31.

111. Matsuda’s system maintains a working memory that stores information about a current conversation with a user. *Id.*, ¶23. For example, the system stores information about a person who is the subject of a conversation in a memory, by setting a “subject person number” in the working memory indicating the person number of the person who is the subject of the current conversation, if any. *Id.*, ¶26. If the subject person number is zero, then there is no current subject person. If the subject person number is at least one, then “the person indicated by that [person] number [is] the subject.” *Id.* By storing information of the conversation subject, the system is able “to progress [the conversation] naturally while returning appropriate responses in relation to the subject.” EX1007, ¶12.

112. When a user first mentions a person during a conversation with the system, Matsuda’s system establishes that person as the current subject by storing

the subject's person number as the subject person number. EX1007, ¶¶26, 39. When subject person number is established, Matsuda's system also tracks the elapsed time since the person indicated by the subject person number in the working memory was last mentioned by the user. *Id.*, ¶¶27 (“[t]he ‘subject person number latest setting date/time’ is used for identifying that the prescribed time has elapsed, indicating the date/time at which the name of the person who is the current subject was last mentioned by the user that is the participant in the conversation”), 69 (recording the current date/time when the subject person number is set). The system expires the subject person number if the elapsed time exceeds “a prescribed amount of time” (“*a predetermined time period*”). EX1007, ¶¶27, 38, 69. Specifically, “if [the] prescribed time (for example, 10 minutes) has elapsed since” the subject person was last mentioned by the user, Matsuda's system resets the subject person number to zero and clears the timer tracking the elapsed time. *Id.*, ¶¶27, 38, 69. This way, the system avoids referencing the wrong person in the conversation. *Id.*, ¶¶95, 100 (“if the user has, for a prescribed period of time, not used the name of a person, the subject person number is set to 0, and no statement is made regarding that person, thus enabling appropriate statements corresponding to the conversation having moved away from the subject of that person.”).

113. Matsuda discloses and renders obvious “*accumulating, by the computer system, short-term knowledge based on one or more natural language*

utterances received during a predetermined time period,” by “*accumulating*” name information about the subject person derived from the “natural language utterances” it receives from the user during the “*predetermined time period*” for a current subject of a conversation. EX1007, ¶¶62-67. A given person may have multiple names, all associated with the same person number in the name data. *Id.*, ¶31 (“The ‘person number’ indicates the number of that person. The ‘name’ indicates the name of that person. In this case, a plurality of names is included for a single individual.”). If the user establishes a conversation subject and then utters a similar name during the “prescribed time” (“*predetermined time period*”), the system treats that similar name as an additional name for the conversation subject by assigning the similar name the same person number in the name data. *Id.*, ¶¶31, 62-66. For example, if there is no current conversation subject, and the user says “I like Matsui of the Kyo**,” the system assigns a person number corresponding to “Matsui of the Kyo**” as the subject person number for the conversation. *Id.*, ¶74. If the user then says “Matsui is strong” within the predetermined time period (i.e., while “Matsui of the Kyo**” is still the current subject), the system determines that “Matsui” is the same person as “Matsui of the Kyo**” and therefore assigns the same person number to “Matsui.” *Id.*, ¶75. Conversely, if the user says “How is Matsui doing?” after the predetermined time period (i.e., after the timer has expired and there is no longer a current subject), Matsuda’s system uses a different set of rules to determine whether

“Matsui” refers to a person the user previously discussed stored in the name data. *See id.*, ¶¶77-80 (describing rules for determining whether the user identified a new subject).

114. Thus, as I describe above, when the user establishes a person as the subject of the conversation, Matsuda’s system keeps this person as the conversation subject “*during a predetermined time period.*” *Id.*, ¶¶26, 27, 38. And the system “*accumulate[es]*” information about the subject “*based on one or more natural language utterances received during [the] predetermined time period,*” e.g., by accumulating information of other similar names such as “Matsui” by which the user refers to “Matsui of the Kyo**,” when spoken during the predetermined time period. *Id.*, ¶¶31, 62-66. A POSITA would have understood that the accumulated information is “*short-term knowledge*” because it is knowledge that “may accumulate during a single conversation, where input received during a single conversation may be retained.” EX1001, 5:10-12, 14:5-6, 16:59-62, 17:26-29.

115. A POSITA would have been motivated to combine Kennewick and Matsuda, and the combination discloses and renders obvious “*accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system.*”

Kennewick and Matsuda are analogous art; both describe dialog-based interaction systems that accumulate short-term knowledge and use the knowledge to provide computer-generated responses to human natural language input. Kennewick's system builds "user dialog history," which is short-term knowledge for a conversation, and uses it to identify meaning and take actions. EX1004, ¶¶155-162. Matsuda's system describes a specific way of accumulating and using short-term knowledge for a conversation during a predetermined time period, e.g., by identifying the current subject of a conversation and detecting additional names for the subject mentioned within a predetermined time period, and using the accumulated information to interpret received user inputs. EX1007, ¶¶26, 75.

116. A POSITA would have found it obvious to incorporate Matsuda's disclosure for identifying a conversation subject and additional names for the subject spoken within a predetermined time period into Kennewick's system that accumulates short-term knowledge based on one or more received natural language utterances related to a single conversation between a user and the computer system. Matsuda discloses that accumulating information about the conversation, e.g., additional names for a conversation subject mentioned within a predetermined time period, "makes it possible to achieve a natural conversation" between the user and system. EX1007, ¶116. Kennewick's system also uses available information, including short-term knowledge, "to achieve a natural environment for one or more

users making queries or commands in multiple domains.” EX1004, ¶11. Accordingly, a POSITA would have been motivated to incorporate Matsuda’s accumulated information into Kennewick’s short-term knowledge to achieve the “natural question and response environment,” by enabling the conversation “to progress naturally while returning appropriate responses in relation to [an identified] subject” of the conversation. EX1004, ¶32, EX1007, ¶12.

117. This is a combination of prior art elements according to known methods to yield the predictable result of improving Kennewick’s natural language interactive speech processing. *See, e.g.*, EX1004, ¶118 (“The natural language interactive speech processing system may make maximum use of context, prior information, location information, domain information and user specific profile data to achieve a natural environment for one or more users making queries or stating commands across multiple domains.”). A POSITA would have had a reasonable expectation of success in making this combination, as both systems are designed to manage conversational flow, operate on similar types of input, and serve the shared goal of producing human-like system responses in natural language dialogs.

[1.f] “accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;”

118. Kennewick in combination with Matsuda discloses and renders obvious limitation [1.f]. As I discuss for limitation [1.e], Kennewick in combination with

Matsuda discloses and renders obvious the “computer system” “accumulating...short-term knowledge based on one or more natural language utterances received during a predetermined time period.” The combination of Kennewick and Matsuda further discloses the “*computer system*” “*accumulating...long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period.*”

119. Kennewick discloses the “*computer system*” “*accumulating...long-term knowledge.*” The ’699 patent describes long-term knowledge as “generally...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.” EX1001, 5:29-33. In addition, “[l]ong-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.” *Id.*, 5:35-39.

120. Kennewick’s computer system “*accumulat[es]*” information from past or historical conversations in a user profile, by “creat[ing], stor[ing] and us[ing] extensive personal profile information for each user.” EX1004, ¶31. The user profile includes, among other things, user information such as the user’s “special use

of terminology,” a “history of questions asked,” “session histories,” a “history of commands and queries,” and a “history of [individual] interactions with the system”. *Id.*, ¶¶31, 143. The historical information in the profile is unconstrained by the current session. *See, e.g., id.*, ¶31 (“Information in the profile may be added and updated automatically as the user uses the system or can be manually added or updated by the user.”). A POSITA would have understood, therefore, that information within the user profile, which is user-centric and unconstrained from the current session, is “*accumulat[ed]...long-term knowledge.*”

121. A POSITA would have understood that Kennewick’s system “*accumulat[es]...long-term knowledge...based on one or more natural language utterances received prior to the*” single conversation of limitation [1.e]. Kennewick discloses that conversations have beginnings and ends. *Id.*, ¶156 (“A dialog with the system may begin when a user first addresses it...At the end of a dialog or to interrupt a dialog, the user may speak a dismissal word (‘good by’).”). Accordingly, a POSITA would have understood and found obvious that a user may have multiple conversations with Kennewick’s computer system, including conversations before (“*prior to*”) and after the single conversation, and that the computer system receives natural language utterances from the user during these conversations. Therefore, a POSITA would have understood and found obvious that the computer system accumulates long-term knowledge based on these conversations, such as the “history

of questions asked,” “session histories,” a “history of commands and queries,” a “history of [individual] interactions with the system,” and information concerning the user’s “special use of terminology.” *Id.*, ¶143.

122. To the extent that Kennewick does not explicitly disclose “*accumulating...long-term knowledge...based on one or more natural language utterances received prior to the predetermined time period.*” Matsuda does.

123. Matsuda discloses “*the predetermined time period*” as I discuss for limitation [1.e]. Matsuda’s system tracks the elapsed time since the subject person corresponding to the subject person number in the working memory was last mentioned by the user, and resets the subject person number to zero if the elapsed time exceeds “*a predetermined time period*” without the subject person being mentioned in a new utterance from the user. EX1007, ¶¶26, 27, 38.

124. Matsuda’s system “*accumulat[es]...long-term knowledge...based on one or more natural language utterances.*” Matsuda’s system persistently stores the name data (e.g., subjects’ names and person numbers) in a nonvolatile memory where “the stored content is not lost even in the event that the power supply is cut,” allowing for information to be accumulated over multiple conversations. *Id.*, ¶¶28, 31. The name data includes information for prior subject persons (i.e., name data for persons who were previously but are no longer the current conversation subject), and Matsuda’s system uses the stored name data to identify subjects referenced in

subsequent conversations. *See id.*, ¶¶77-84 (describing using information from prior conversations to identify current conversation subjects). For example, the system may use stored name data relating to how a user refers to “Matsui of the Kyo**” accumulated from prior conversations to identify a conversation subject of a later conversation. *Id.*

125. A POSITA would have understood that the name data is “*accumulat[ed]...based on one or more natural language utterances*” as users converse with the system over time. A POSITA also would have understood that the persistently stored name data is “*long-term knowledge*” because it is user-centric and persists across multiple conversations. *Id.*, ¶¶28, 31 (user-centric name data is stored in nonvolatile memory); *see* EX1001, 5:29-39 (“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 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.”).

126. A POSITA would have also understood that Matsuda discloses “*accumulating...long-term knowledge...based on one or more natural language utterances received prior to the predetermined time period.*” The ’699 patent

discloses that “relevant information from an expired session context may nonetheless be added to ... long-term knowledge models.” EX1001, 14:39-42. Likewise, Matsuda also discloses that the name data persistently stored in the nonvolatile memory includes information for prior subject persons (i.e., expired subjects that are no longer the current conversation subject). EX1007, ¶¶77-84. For example, Matsuda’s system accumulates long-term knowledge based on natural language utterances received during prior conversations. *See, e.g., id.*, ¶¶76-77 (“Let us assume that the user next says ‘Matsui is good at defense as well,’ ... Let us assume that the conversation on this date is terminated in this state and, on a later date, the user suddenly asks, ‘How is Matsui doing?’”). A POSITA thus would have understood that at least some of the name data is from conversations that occurred “*prior to a predetermined time period*” of the current conversation, i.e. before the user established the predetermined time period by introducing a conversation subject into the current conversation. *See, e.g., id.*, ¶77 (“Let us assume that the conversation on this date is terminated in this state and, on a later date, the user suddenly asks, ‘How is Matsui doing?’ At this time, the ‘subject person number’ is 0”). Therefore, Matsuda’s long-term knowledge is obviously accumulated “*based on one or more natural language utterances received prior to the predetermined time period.*”

127. A POSITA would have been motivated to combine Kennewick and Matsuda, and the combination discloses and renders obvious “*accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period.*” Kennewick and Matsuda are analogous art; both describe dialog-based interaction systems that accumulate long-term knowledge and use the knowledge to provide computer-generated responses to human natural language input. *See, e.g.*, EX1004, ¶11 (“This environment makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands in multiple domains.”); EX1007, ¶84 (“In this way, when, in this system, a given person is to be the subject, it is not necessary always to use the exact same name, making it possible to have a conversation with the name that is used without having to think about it, as is the case in a conversation held between people.”).

128. A POSITA would have found it obvious to incorporate Matsuda’s disclosure for identifying a conversation subject using accumulated *long*-term knowledge into Kennewick’s system. Matsuda discloses that persistently storing information based on natural language utterances received prior to a predetermined time period, such as name data, “makes it possible to achieve a natural conversation” between the user and system, and allows the system to account for the individual

vocabulary and speaking habits of different users. EX1007, ¶¶101 (storing persistent name data “enables each individual conversation participant to use...the names that he or she always uses”), 116. Kennewick’s system “makes significant use of context, prior information, domain knowledge, and user specific profile data,” such as user preferences and the user’s “special use of terminology,” “to achieve a natural environment for one or more users making queries or commands in multiple domains.” EX1004, ¶11. Accordingly, a POSITA would have been motivated to incorporate Matsuda’s disclosure related to accumulating and using information received prior to a predetermined time period, e.g., name data, to provide additional information for use by Kennewick’s system to identify conversation subjects and achieve the natural environment. A POSITA would have likewise found it obvious for other data accumulated in Kennewick’s user specific profile, e.g., the user’s “special use of terminology,” “history of questions asked” and “interactions with the system,” etc., to include information accumulated based on natural language utterances received prior to a “predetermined time period,” as taught by Matsuda. *See also* EX1007, ¶¶108, 113 (embodiments set forth in Matsuda are not limited to person names).

129. The combination of Kennewick and Matsuda is a combination of prior art elements according to known methods to yield predictable results. The combination of Kennewick and Matsuda is also an application of a known

technique – Matsuda’s technique for accumulating and using long-term knowledge – to a known system – Kennewick’s speech-based conversational interface that uses long-term knowledge including user profile information – ready for improvement. A POSITA would have found the results of this combination predictable and would have reasonably expected the results to be successful, because it would enable the combined system to better respond to the user in a more natural way.

[1.g] “identifying, by the computer system, a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and”

130. Kennewick in combination with Cooper discloses and renders obvious limitation [1.g]. Kennewick discloses receiving a user input comprising a natural language utterance as I discuss for limitation [1.a]. Kennewick and Matsuda further disclose accumulating “*short-term knowledge*” and “*long-term knowledge*” as I discuss for limitations [1.e] and [1.f].

131. Further, Kennewick discloses using different personalities for responses. Kennewick explains that “[e]xamples of personality traits” that may be included in its responses “include, sarcasm, humor, irritation, and sympathy, and other traits.” EX1004, ¶141. The use of personality traits when generating responses enables Kennewick’s system to respond appropriately to the user, such as presenting

upsetting information “in a sympathetic manner” or urgent alerts “with a definite and authoritative personality.” *Id.*, ¶24.

132. However, Kennewick does not explicitly disclose that its computer system “*identiffies]...a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge.*” This aspect is taught by Cooper.

133. Cooper describes a “virtual assistant [that] adapts to the user in many different ways based on the input the virtual assistant receives.” EX1005, 2:56-64, Abstract. The input includes speech (natural language utterance) received from a user. EX1005, Abstract (disclosing “a voice user interface for inputting information into . . . the virtual assistant by speech”), 6:35-37 (“Using speech recognition and text-to-speech technology, the VA [virtual assistant] communicates with callers in spoken English.”), 7:50-53 (“the VA Engine uses its telephony interface 104 to communicate with the user 106 and its speech interface 110 to recognize speech into text and translate text into speech”), 46:58-59 (“in making utterances to the VA, the user should speak clearly and at a moderate pace”).

134. Cooper’s virtual assistant “*identiffies]...a manner in which the natural language utterance was spoken.*” The virtual assistant determines “whether words associated with polite discourse are included in input received from the user.” EX1005, 43:55-59, 2:26-30 (criticizing prior art’s failure to automatically adapt

when the “user’s input contains words associated with polite discourse”). Cooper also explains that “the user information input into the virtual assistant could be information about the user’s emotion, which could be based on information about the user’s voice volume, word choice and speech rate. Based on such information, the virtual assistant could automatically determine the user’s emotional state, calm or angry, for example.” *Id.*, 43:62-67. Cooper thus “*identifies]...a manner in which the natural language utterance was spoken*” by determining whether the utterance included words associated with polite discourse (*id.*, 43:55-59) and by determining whether the user’s word choice conveys an emotional state such as being calm or angry (*id.*, 43:62-67).

135. Further, a POSITA would have found it obvious to combine Cooper with Kennewick such that Kennewick’s computer system identifies “*a manner in which the natural language utterance was spoken based on the short-term knowledge.*” Kennewick’s computer system accumulates the short-term knowledge from the dialog with the user. EX1004, ¶¶23, 25, 29, 31, 37, 155, 157, 160. Cooper discloses using the words of the dialog to identify the manner in which the user is speaking. EX1005, 43:55-59, 44:1-4. Accordingly, a POSITA would have found it obvious in Kennewick’s computer system modified by Cooper to identify the manner in which the natural language utterance was spoken based on the short-term knowledge, because the short-term knowledge includes information (words or

phrases) about the dialog in which the user made the utterance, e.g., words or phrases tagged to enable “other components of the system to tie that utterance to the correct user and dialog.” EX1004, ¶155.

136. A POSITA also would have found it obvious to combine Cooper with Kennewick such that Kennewick’s computer system identifies “*a manner in which the natural language utterance was spoken based on the long-term knowledge.*” Kennewick’s computer system accumulates long-term knowledge by storing extensive personal profile information for each user including information from past sessions between the user and the computer system, including, among other things, a “history of questions asked,” “session histories,” a “history of commands and queries,” and a “history of [individual] interactions with the system”. EX1004, ¶¶31, 143. Cooper also saves “*long-term knowledge*” corresponding to information from past sessions, including information obtained about the user’s emotional state such as word choice information, “for use in future sessions.” EX1005, 43:62-44-6. For example, Cooper may compare a user’s saved emotional state information of a prior session (e.g., user’s word choice indicating a particular emotional state in a prior session) with the user’s word choice in a current session to identify the user’s manner of speaking (i.e., emotional state) in the current session (i.e., similar word choice may indicate a similar emotional state across sessions). *Id.*, 43:55-44-6. Cooper thus discloses using long-term knowledge, e.g., emotional state information from

past sessions, to identify a manner of the user's natural language utterance, e.g., the user's emotional state when making the utterance, in a later session. *Id.* A POSITA would have found it obvious to modify Kennewick's computer system based on Cooper to accumulate information of the user's emotional state from past sessions, such as word choice information, as part of the accumulated long-term knowledge. The system can then use this information to identify the manner in which the natural language utterance was spoken as taught by Cooper, because the long-term knowledge includes information from past sessions based on which the user's manner of speaking for a current session can be identified. EX1004, ¶¶31, 143; EX1005, 43:62-44-6.

137. Still further, the combination of Kennewick and Cooper discloses and renders obvious “*identifying, by the computer system, a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge.*” As I discuss above, Cooper identifies a user's manner of speaking based on the user's word choice in an utterance in the current conversation (short-term knowledge) and saved emotional state information from past conversations (long-term knowledge), to identify the user's particular manner of discourse, such as polite discourse and/or discourse corresponding to a particular emotional state, e.g., angry, calm, etc. EX1005, 43:55-44:6. For example, by identifying the user's manner of speaking for a current session based on whether

words included in the user's utterances in the current session (short-term knowledge) are similar to saved emotional state information (long-term knowledge) including words in the user's utterances in past sessions, Cooper's system determines the user's manner of speaking using both short-term and long-term knowledge. EX1005, 43:55-44:6. Kennewick's system "makes significant use of context, prior information, domain knowledge, and user specific profile data," which includes short- and long-term knowledge, "to achieve a natural environment for one or more users making queries or commands in multiple domains." EX1004, ¶11; *see also id.*, ¶178 ("domain agent may...produce a response best understood and most natural to the user."). A POSITA also would have found it obvious to combine Cooper with Kennewick such that Kennewick's computer system, as part of making use of "context, prior information,...user specific profile data," etc., uses both the words in the utterance in the current session (short-term knowledge) and the saved emotional state information including words in utterances in past sessions (long-term knowledge) to identify the manner in which the utterance was spoken as taught by Cooper. Using information from both current and past sessions to identify the manner would further Kennewick's stated goals of achieving a natural environment and producing a response that is best understood and most natural to the user.

138. This is a combination of prior art elements according to known methods to yield a predictable result of identifying a manner in which an utterance was spoken

based on short-term and long-term knowledge about the user. The combination is also a predictable use of known techniques for improving voice-based interactions with computers. The improvement would have been within the skill of a POSITA, who would reasonably expect success based on the compatibility of the systems and their shared goal of enabling adaptive, conversational interfaces.

[1.h] “generating, by the computer system, a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.”

139. Kennewick in combination with Cooper discloses and renders obvious limitation [1.h]. As I discuss for limitation [1.d], Kennewick’s computer system determines the “*interpretation*” of the natural language utterance. EX1004 ¶¶123, 155, 160, 162-63, 165. Further, as I discuss for limitation [1.g], Kennewick in combination with Cooper discloses identifying the “*manner in which the natural language utterance was spoken.*” EX1005, 43:51-44:6; 2:26-30. As I discuss for the preamble [1.pre], Kennewick in combination with Cooper also discloses that generated responses are “*adapted based on a user’s manner of speaking.*”

140. Kennewick further discloses and renders obvious its computer system “*generating [] a response to the natural language utterance based on the interpretation.*” See, e.g., EX1004, ¶14 (“the system...receives spoken natural language queries, commands and/or other utterances from a user,...processes the input..., and responds to the user with a natural language speech response.”).

141. Kennewick's computer system invokes an agent that works with the parser to interpret the utterance based on the determined context. *Id.*, ¶¶162 (“[o]nce the context for the question or command has been determined, the parser 118 can invoke the correct agent”), 163 (describing the “parser and [the invoked] domain agent” using “context to determine the most plausible interpretation”). The agent then generates an appropriate response to user's natural language utterance for output in a suitable presentation format. *Id.*, ¶¶17 (“[a]gents are executables that receive, process and respond to user questions, queries and commands”), 124 (“[t]he agents 106 may return results of questions as responses to users”), 32 (“[t]he agent may make one or more queries and may rapidly return a formatted response”), 178 (“[o]nce the domain agent 156 has created a satisfactory response to a question, or to a command, the agent may format that response for presentation”). And “[t]ypically, the domain agent can format the response into the markup format used by the text to speech engine 124.” *Id.*, ¶¶178, 124 (“[t]he agent may create a response string, which may be sent to the text to speech engine 124.”) The “text to speech engine creates the required utterances,” which after further processing are “output by the speaker 136.” *Id.*, ¶¶124, 23-25, 32-33, 178-82 (describing, among other things, templates, transformations, other display formats that can influence presentation of the output).

142. Given Kennewick’s description of system components that generate responses to user utterances, a POSITA would have understood Kennewick to disclose and render obvious “*generating, by the computer system, a response to the natural language utterance based on the interpretation.*” A POSITA would have understood and found obvious that the response is based on the interpretation of the natural language utterance because Kennewick’s parser and agent generate the interpretation, which the agent uses to generate the response. *See, e.g.,* EX1004, ¶¶28, 162-163, 178-179.

143. Further, Kennewick provides for adapting the response format based on a “simulated personality.” EX1004, ¶¶24, 178-180. That is, Kennewick discloses generating responses based on more than just semantic content, but also based on the manner in which the responses are to be conveyed to the user (e.g., sympathetic, authoritative, etc.). *Id.* However, Kennewick does not explicitly disclose formatting the response based on the manner in which the utterance was spoken. This aspect is taught by Cooper.

144. Cooper teaches generating the response based on not only the interpretation of the natural language utterance but also “*the identified manner in which the natural language utterance was spoken.*” For example, Cooper discloses that if “words of polite discourse are included in the input received from the user, the virtual assistant could automatically enable the politeness setting.” EX1005,

43:59-61. And the “politeness setting . . . when enabled, would cause the virtual assistant to include words or phrases associated with polite discourse in the output from the virtual assistant.” *Id.*, 43:48-51. As another example, Cooper explains that “[i]f the user’s emotional state is angry, the output of the virtual assistant could automatically include words associated with submissive discourse, such as ‘sorry,’ ‘regret’ and ‘apologize.’” *Id.*, 43:62-44:4. The disclosure in Cooper parallels the Adaptive Response Builder 315 in the ’699 patent, which “may create a conversational feel by adapting to a user’s manner of speaking...by varying tone, pace, timing, inflection, word use, jargon, and other variables in a verbal or audible response.” EX1001, 18:45-50; *see also id.*, 18:65-19:19 (showing “how a response may adapt to a user’s input word choices and manner of speaking”).

145. A POSITA would have found it obvious to incorporate Cooper’s teachings regarding responding appropriately to the user’s manner of speech (e.g., incorporating words or phrases of polite or submissive discourse in the output when the user’s input includes polite discourse or indicates an angry emotional state) into the agent-based response system of Kennewick because both references address improving the naturalness and user-friendliness of voice-based interactions. Kennewick describes agents that “create a response string” and “format that response for presentation,” (EX1004, ¶¶124, 178), e.g., to include “simulated personality characteristics” selected based upon various factors (*id.*, ¶¶24, 35), while

Cooper discloses selecting different patterns of words based on user input manner to enhance the interaction (EX1005, 43:48–44:4), e.g., to complement the user’s politeness level and/or emotional state. Combining Cooper’s method of selecting a polite or submissive tone with Kennewick’s system, so that Kennewick selects a “system personality” for formatting responses based on the identified manner of the user’s input as taught by Cooper, would have been a straightforward application of known techniques to improve the user experience. *See* EX1004, ¶¶24, 178.

146. Such a combination would yield predictable results. Kennewick already discloses that formatting responses to “simulate[] personality characteristics” is desirable (EX1004 ¶24), while Cooper teaches specific ways to generate responses based on the user’s identified manner (EX1005, 43:55–44:4), and it would have been obvious to modify Kennewick’s system to generate and present responses in a polite manner responsive to the user input being indicative of polite discourse as taught by Cooper. A POSITA would have reasonably expected the resulting system to successfully generate responses not only based on the semantic interpretation of an utterance but also on how (i.e., manner in which) that utterance was spoken, because these technologies were well-understood.

147. Therefore, the combination of Kennewick and Cooper discloses and renders obvious “*generating, by the computer system, a response to the natural*

language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.”

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

148. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 (“systems and methods are provided”). Further, as I discuss for [1.pre] and [1.g], Cooper discloses identifying the “*manner in which the natural language utterance was spoken,*” such as by determining that the user’s input includes “words associated with polite discourse” or contains information indicating the user is “calm or angry.” EX1005, 43:55-67.

149. Cooper further discloses and renders obvious that the identified manner “*includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.*” For example, Cooper discloses determining whether the “input received from the user” contains “words associated with polite discourse.” EX1005, 43:55-61. In addition, Cooper also discloses that “the user information input into the virtual assistant could be information about the user’s emotion, which could be based on information about the user’s voice volume, word choice and speech rate. Based on such information, the virtual assistant could automatically determine the user’s emotional state, calm or angry, for example.” EX1005, 43:62-67. Further, as

I discuss for preamble [1.pre], Cooper's virtual assistant adapts the words in the output response based on word use detected for the received input. *Id.*, 43:49-61.

150. A POSITA would have understood that Cooper's virtual assistant determining "whether words associated with polite discourse are included in input received from the user," or detecting "information about the user's voice volume, word choice and speech rate" of the received input amounts to Cooper's virtual assistant identifying "*an indication of at least one of ... word use, and/or jargon,*" in the received input corresponding to the manner. A POSITA would have further understood that Cooper's virtual assistant detecting "information about the user's ... speech rate" in the received input amounts to Cooper's virtual assistant identifying "*an indication of at least one of ... pace, [and/or] timing*" in the received input corresponding to the manner.

151. Therefore, Cooper discloses and renders obvious "*wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.*"

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

152. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 ("systems and methods are provided").

153. Kennewick further discloses “*wherein the response comprises a voice response.*” Based on an interpretation of the user’s natural language utterance, Kennewick’s computer system invokes an agent, which in turn generates and outputs an appropriate response to user. EX1004, ¶¶124, 162. Kennewick further discloses sending the response to a text to speech engine which creates the utterances and outputs them from a speaker, which a POSITA would have understood to comprise “*a voice response.*” *Id.*, ¶124 (the “text to speech engine creates the required utterances,” which are “output by the speaker 136.”).

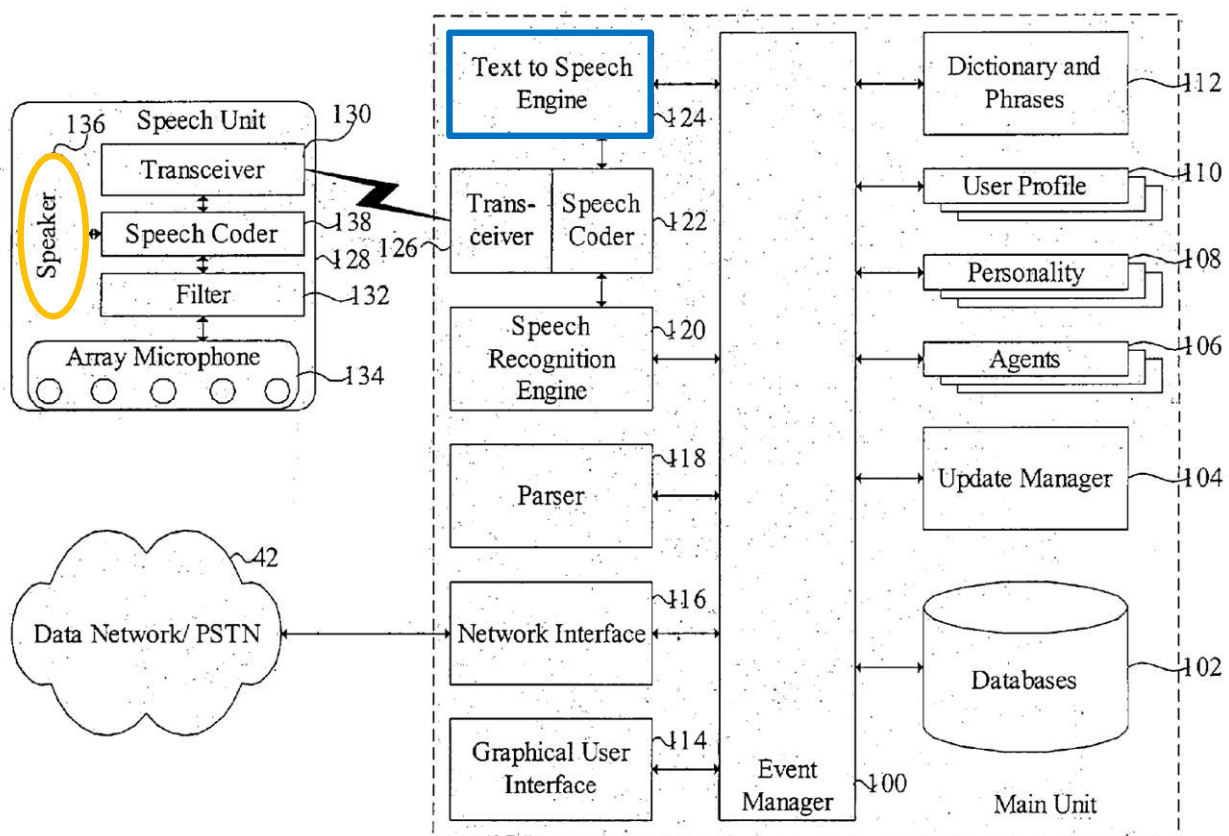


Figure 5. Speech Processing System Block Diagram

EX1004, Figure 5.

154. Further, as I discuss for [1.h], Kennewick in combination with Cooper discloses and renders obvious generating the “*response based on the identified manner.*” Since Kennewick further discloses that the response may be a voice response, a POSITA would have understood that Kennewick’s computer system, as modified by Cooper’s manner-based response adaptation logic, discloses and renders obvious “*generating the voice response based on the identified manner.*”

155. Cooper further discloses its virtual assistant adapts to the user’s manner of speaking by “*varying one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.*”

156. Like Kennewick, Cooper explicitly discloses the response generated by its virtual assistant based on the identified manner may be a “*voice response.*” Cooper’s virtual assistant communicates with callers using “text-to-speech technology ... in spoken English.” EX1005, 6:35-37; *see also id.*, 8:5-7 (“Text-to-Speech Server 120 receives text from other components, translates it into speech (that is, into a sound file), and returns it to the requesting component.”).

157. Further, Cooper discloses that its virtual assistant “*var[ies] one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response*” when generating the voice response based on the identified manner. EX1005, 43:48-61 (discussing including words or phrases associated with polite discourse in the output from the virtual assistant when the politeness setting is enabled); *see also*

id., 43:62-44:4 (discussing including words associated with submissive discourse in the output when the user’s emotional state is detected as angry).

158. A POSITA would have understood that Cooper’s virtual assistant (1) automatically including words or phrases associated with polite discourse in the voice response output from the virtual assistant in response to recognizing words associated with or containing polite discourse in the speech input received from the user; and/or (2) automatically including words associated with submissive discourse in the voice response output from the virtual assistant in response to determining a user’s emotional state (e.g., angry) in the speech input received from the user, meets the claimed “*varying one or more of ... word use, and/or jargon of the voice response.*”

159. A POSITA would have found it obvious to incorporate Cooper’s teaching of varying word use or jargon of the voice response based on the user’s manner of speaking into Kennewick’s system because doing so is consistent with Kennewick’s stated objectives of creating “a complete speech-based natural language query and response environment” able to “simulate some aspects of human ‘personality.’” EX1004, Abstract, ¶¶11, 24. By varying the word use and/or jargon of the presented responses based on the identified manner as taught by Cooper, the system can better “adapt its behavior...to the user... based on the input the virtual

assistant receives” from the user, and present responses that appear less “rigidly formatted or mechanical.” EX1004, ¶24; EX1005, Abstract.

4. Claim 4

[4.pre] “The method of claim 1, the method further comprising:”

160. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 (“systems and methods are provided”).

[4.a] “obtaining, by the computer system, contextual signifiers and/or grammatical rules,”

161. A POSITA would have found it obvious that Kennewick’s computer system “*obtain[s]...grammatical rules*” to format responses to user utterances. More specifically, Kennewick’s computer system invokes an agent used to generate an appropriate response to the user’s natural language utterance. EX1004, ¶¶17, 124, 132, 162. Kennewick further discloses that after creating a satisfactory response, “the agent may format that response for presentation.” *Id.*, ¶178.

162. A POSITA would have found it obvious that Kennewick’s computer system formatting a response to a user’s utterance would involve “*obtaining, by the computer system,...grammatical rules*” to format the response. For example, Kennewick discloses using “presentation templates” to format responses “in a manner enhancing the understandability to the user,” so that the responses can be communicated to the user as “natural language speech response[s].” EX1004, ¶¶14, 23, 179, 181. A POSITA would have understood that a “template” that specifies the

format of a “natural language speech response” would specify a set of rules or guidelines how to generate the response. *See also id.*, [0181] (“template” specifies how to “summariz[e] the results” to be presented to the user). For a speech response, such rules would include those relating to the grammar and syntax of the response, so that the generated speech response will be grammatically correct and thus understandable to the user. *See id.* Furthermore, a POSITA would have found it obvious for the formatting of response to involve the use of “grammatical rules,” because this was how conventional dialog manager systems operated. For example, as discussed in the well-known Huang textbook, conventional dialog manager systems obtain grammars and rules when generating (formatting) responses to queries. *See, e.g.*, EX1011, 912 (discussing a dialog manager that “generates the desired answer for the query.”), 913 (discussing using a dialog grammar whose “rules state sequential and hierarchical constraints on acceptable dialogs, just as syntactic grammar rules state constraints on grammatically acceptable strings.”); *see also id.*, 924 (Figure 17.26 showing formatting of the semantic response from the dialog manager by obtaining a “Grammar” and “Rules for Speech Synthesis” and applying the obtained “Grammar” and “Rules for Speech Synthesis” on the semantic response from the dialog manager to generate a synthetic speech response).

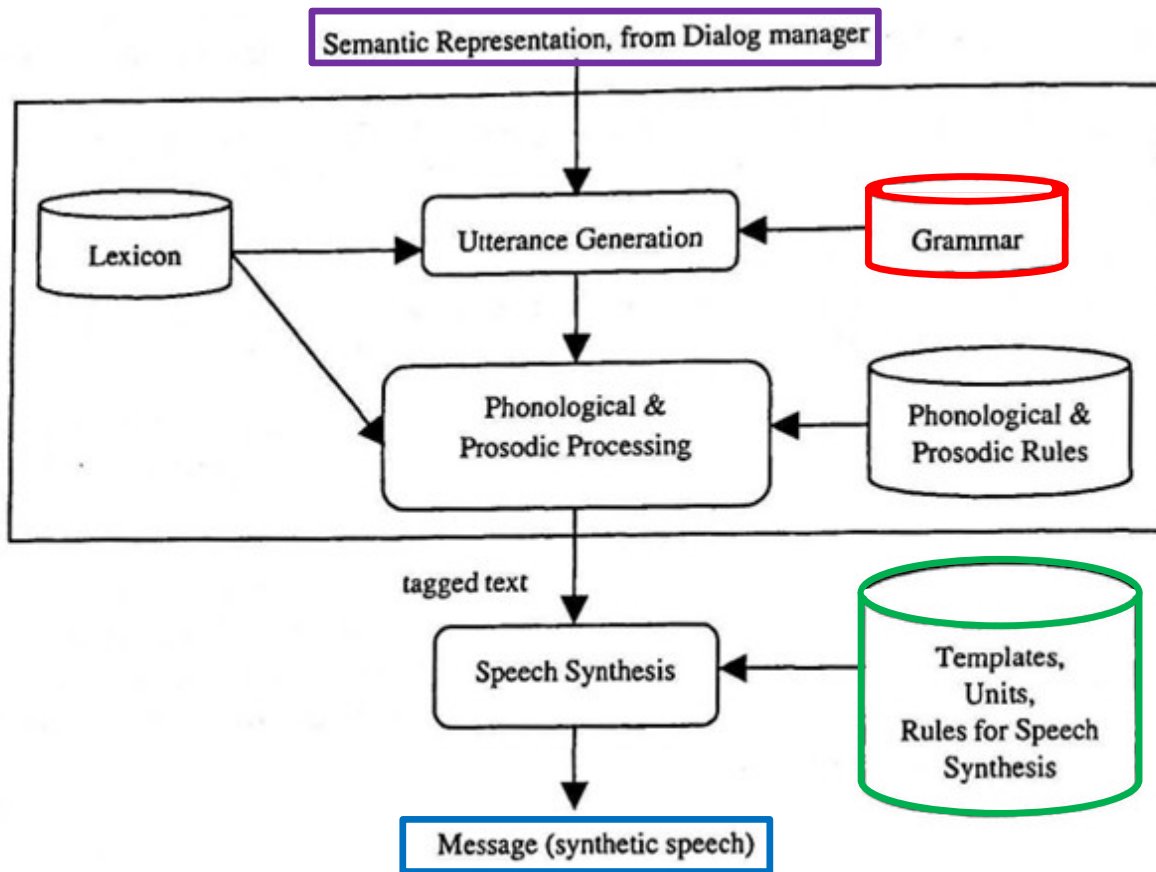


Figure 17.26 Natural language generation and rendition modules.

163. A POSITA would thus have found it obvious that Kennewick’s computer system formatting a response to a user’s utterance would involve *“obtaining, by the computer system, ... grammatical rules”* to format the response to generate a “natural language speech response” that is grammatically correct and is understandable to the user.

[4.b] “wherein generating the response based on the identified manner in which the natural language utterance was spoken comprises using the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.”

164. As I discuss for limitation [1.h], Kennewick in combination with Cooper discloses and renders obvious ***“generating the response based on the identified manner in which the natural language utterance was spoken.”*** Further, as I discuss for limitation [4.a], Kennewick renders obvious ***“obtaining, by the computer system, ... grammatical rules.”***

165. Kennewick also discloses and renders obvious ***“using the obtained ... grammatical rules to generate sentences for use as response sets to cooperate with the user.”*** First, a POSITA would have understood that Kennewick’s computer system ***“obtaining...grammatical rules”*** to format responses would involve ***“using”*** the obtained grammatical rules to generate the formatted response. *See supra* [4.a]. Kennewick further discloses that the response may be a “natural language speech response,” which is formatted so it is natural to the user. EX1004, ¶¶14, 178 (discussing performing “transformations to produce a response best understood and most natural to the user”). A POSITA would have also found it obvious that the generated “natural language speech response” includes ***“sentences”*** that follow the obtained grammatical rules, so that the response can be conveyed to the user in a “natural language.” *See* EX1004, ¶¶17, 36. In addition, in cases where the information to be conveyed in the response requires more than a single sentence, a

POSITA would have found it obvious for the system to generate multiple sentences for the response. *See* EX1004, ¶¶17, 36 (“invention may use special procedures to present information that is in the form of long text strings”).

166. Kennewick further discloses using “*response sets*” to present information to the user. *See* EX1004, ¶¶25, 36. For example, the system may divide a “long set of information” into subsets that are presented to the user through a set of one or more responses. EX1004, ¶¶14, 25, 36 (describing that certain subsets may be presented “first”). A POSITA would have found it obvious for the system to treat the multiple sentences generated by the system as response sets, because the system communicates the generated sentences to the user through a set of responses, e.g., “natural language speech response[s].” *Id.*

167. Kennewick further discloses that the sentences generated by the system and used for response sets “*cooperate with the user.*” For example, Kennewick discloses that an objective of the system is to generate and provide “a satisfactory response to [the user’s] question [or] command.” *See, e.g.,* EX1004, ¶¶31-32, 170-178. The generates sentences used as response sets cooperate with the user by “present[ing] information” that answers the user’s questions. *See id.*, ¶36. In addition, a POSITA would have understood that Kennewick’s computer system formatting the response so that it is more natural to the user meets the claimed feature of generating the response “*to cooperate with the user*” because a response that has

been transformed to make it more understandable and natural to the user is a response that cooperates with the user. Therefore, the combination of Kennewick and Cooper discloses and renders obvious “*generating the response based on the identified manner in which the natural language utterance was spoken comprises using the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.*”

5. Claim 5

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

168. Claim 1 discloses a method as I discuss for [1.pre]. See, e.g., EX1004, ¶11 (“systems and methods are provided”). As I discuss for [1.f], Kennewick discloses accumulating “*long-term knowledge*” in a user profile.

169. Kennewick further discloses the long-term knowledge in the user profile “*is associated with a first user.*” Kennewick explicitly discloses that the information in each user profile, which includes the long-term knowledge, is specific to an individual user. EX1004, ¶¶31, 143 (“personal profiles may include information specific to the individual, their interests, their special use of terminology, the history of their interactions with the system, and domains of interest.”). Therefore, a POSITA would have understood that the long-term knowledge in the profile “*is associated with a [] user.*” *Id.*, ¶¶11, 23. It would have

been obvious to a POSITA that the “user” is “*a first user*” because the computer system of Kennewick is designed to work with “one or more users.” *Id.*

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

170. Kennewick discloses limitation [5.a]. Kennewick’s computer system creates and stores for each user, a user profile (“*generating...a profile*”) by accumulating “long-term knowledge” about the user (“*first user*”), such as the user’s “history of questions asked, session histories,” “history of [] interactions with the system,” “history of commands and queries,” etc., and adding it to the profile. EX1004, ¶31 (“creates, stores and uses extensive personal profile information for each user.”), 143. Once the user profile has been created, additional information “may be added” and the profile “updated automatically as the user uses the system.” *Id.* A POSITA would have understood that Kennewick’s user profile 110 which is created for each user and which is automatically updated with long-term knowledge information, e.g., histories, over time as the user uses the system meets the claimed “*generating, by the computer system, a profile associated with the first user based on the long-term knowledge.*” This understanding is consistent with the ’699 patent’s description of the claimed “*profile*” because the ’699 patent discloses that “long-term knowledge may accumulate over time to build user profiles.” EX1001, 14:6-7.

171. Further, as I discuss for limitation [1.c], Kennewick discloses identifying the “*context for the natural language utterance.*” Kennewick also explicitly discloses that its computer system determines the context based on the user profile. *Id.*, ¶160 (“The parser may determine a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents”). A POSITA would have understood this disclosure in Kennewick to meet the claimed limitation “*wherein the context for the natural language utterance is determined based further on the profile associated with the first user.*”

6. Claim 6

[6.pre] “The method of claim 1, wherein generating the response based on the identified manner comprises:”

172. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 (“systems and methods are provided”). As I discuss for [1.h], Kennewick in combination with Cooper disclose the limitation “*generating the response based on the identified manner.*”

[6.a] “adapting, by the computer system, the response based on a response format associated with the identified manner.”

173. Kennewick discloses formatting (adapting) a response based on templates, transformations, and the like. EX1004, ¶¶178 (after creating a satisfactory response “the agent may format that response for presentation...using available format templates.”); *see also id.*, ¶179 (“the domain agent 156 may select

the presentation template, determine order of presentation for tokens and determine variable substitutions and transformations using probabilistic or fuzzy set decision methods.”). A POSITA would have understood that Kennewick’s computer system selecting a presentation template or determining an order of presentation of tokens (i.e., words or phrases of a response), or determining substitutions or transformations to a response, amounts to “*adapting...the response.*”

174. However, Kennewick does not explicitly disclose adapting the response “*based on a response format associated with the identified manner.*” This aspect is taught by Cooper.

175. As I discuss for element [1.g], Cooper discloses adapting the response based on the user’s politeness or emotional state by incorporating specific words (response format) associated with polite or submissive discourse (identified manner) in the response. EX1005, 43:51-55, 44:1-4. A POSITA would have understood that Cooper adapting the response to (1) include words or phrases (response format) associated with polite discourse when politeness (identified manner) is detected in the speech input received from the user; and/or (2) include words or phrases associated with submissive discourse when anger is detected in the speech input received from the user, amounts to adapting the response based on “*a response format associated with the identified manner.*” This is because Cooper adapts the

response by using a particular response format (e.g., words associated with polite or submissive discourse) selected based on the user's identified manner.

176. A POSITA would have found it obvious to incorporate Cooper's teachings of different response formats associated with identified manners into Kennewick's system. Kennewick already discloses adapting the response based on various factors. EX1004, ¶¶179-180. Cooper discloses specific ways for adapting the response, e.g., by using a response format associated with the user's identified manner. EX1005, 43:55-67, 2:26-30. A POSITA would have been motivated to combine the references in order to improve the naturalness and user-friendliness of voice-based interactions and for the other responses provided for limitation [1.f], as the combination would allow Kennewick's system to format its responses to simulate different personality types, based on the identified manner as taught by Cooper. *See, e.g.*, EX1004, ¶¶24 ("The use of simulated personality characteristics is... desirable"), 180.

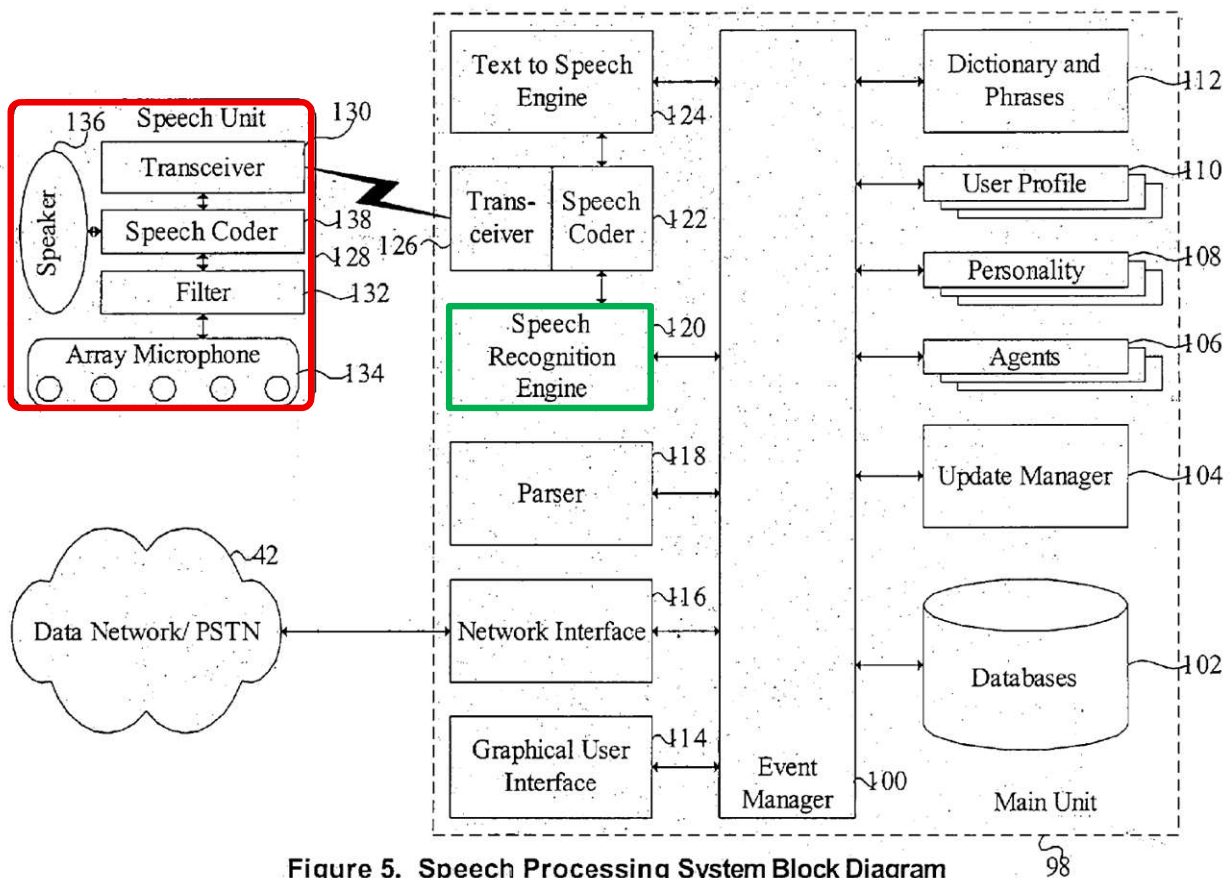
7. Claim 7

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

177. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 ("systems and methods are provided"). As I discuss for [1.b], Kennewick discloses the limitation "*recognizing the one or more words or phrases from the natural language utterance.*"

[7.a] “providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and”

178. Kennewick further discloses limitation [7.a]. Kennewick explicitly discloses a **speech recognition engine** in Figure 5.



EX1004, Figure 5.

179. As I discuss for limitation [1.b], the user’s natural language utterances are received by Kennewick’s **speech unit 128** via the microphone array 134, coded by the speech coder 138 and transmitted to the main unit 98 by the transceiver 130. *Id.*, ¶121. Kennewick further discloses that “[c]oded speech received at the transceiver 126 on the main unit 98 may be passed to the speech coder 122 for

decoding and decompression.” *Id.*, ¶123. The “output of the speech coder 122 may be fed to the **speech recognition engine 120.**” *Id.*, ¶155. A POSITA would have understood that the user’s utterances received by the speech unit 128 and fed to the speech recognition engine 120 in Kennewick meets the claimed limitation **“providing, by the computer system, the natural language utterance as an input to a speech recognition engine.”**

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

180. Kennewick further discloses that the speech recognition engine 120 outputs recognized words and phrases and passes these to the parser 118. Kennewick’s “speech recognition engine 120 may recognize words and phrases, using for example, information in the dictionary and phrase tables 112, and pass these to the parser 118 for interpretation.” *Id.*, ¶155; *see also id.*, ¶123 (“decoded speech can be processed by the speech recognition engine 120 using data in the dictionary and phrases module 112 and received from the agents 106. The recognized words and phrases may be processed by the parser 118, which transforms them into complete commands and questions using data supplied by the agents.”).

181. A POSITA would have understood that Kennewick’s parser 118 receiving the “words and phrases” recognized by the speech recognition engine 120 meets the claimed **“obtaining, by the computer system, the one or more words or**

phrases recognized from the natural language utterance as an output of the speech recognition engine,” because the parser 118 obtains the recognized words and phrases from the speech recognition engine 120.

8. Claim 8

[8.pre] “The method of claim 1, the method further comprising:”

182. Claim 1 discloses a method as I discuss for [1.pre]. See, e.g., EX1004, ¶11 (“systems and methods are provided”).

[8.a] “causing, by the computer system, the response to the natural language utterance to be provided to the user.”

183. As I discuss for limitation [1.h], Kennewick in combination with Cooper discloses generating the *“response to the natural language utterance.”*

184. Kennewick further discloses that its computer system causes the generated response *“to be provided to the user.”* Kennewick’s computer system “create[s] a response string, which may be sent to the [text to speech engine 124](#)” such that the generated response is provided to the user as an “output by the [speaker 136](#).” *Id.*, ¶124, Figure 3. A POSITA would have understood that Kennewick’s computer system outputting the response from the speaker amounts to providing the response to the user because the user of Kennewick’s computer system may hear the response output from the speaker.

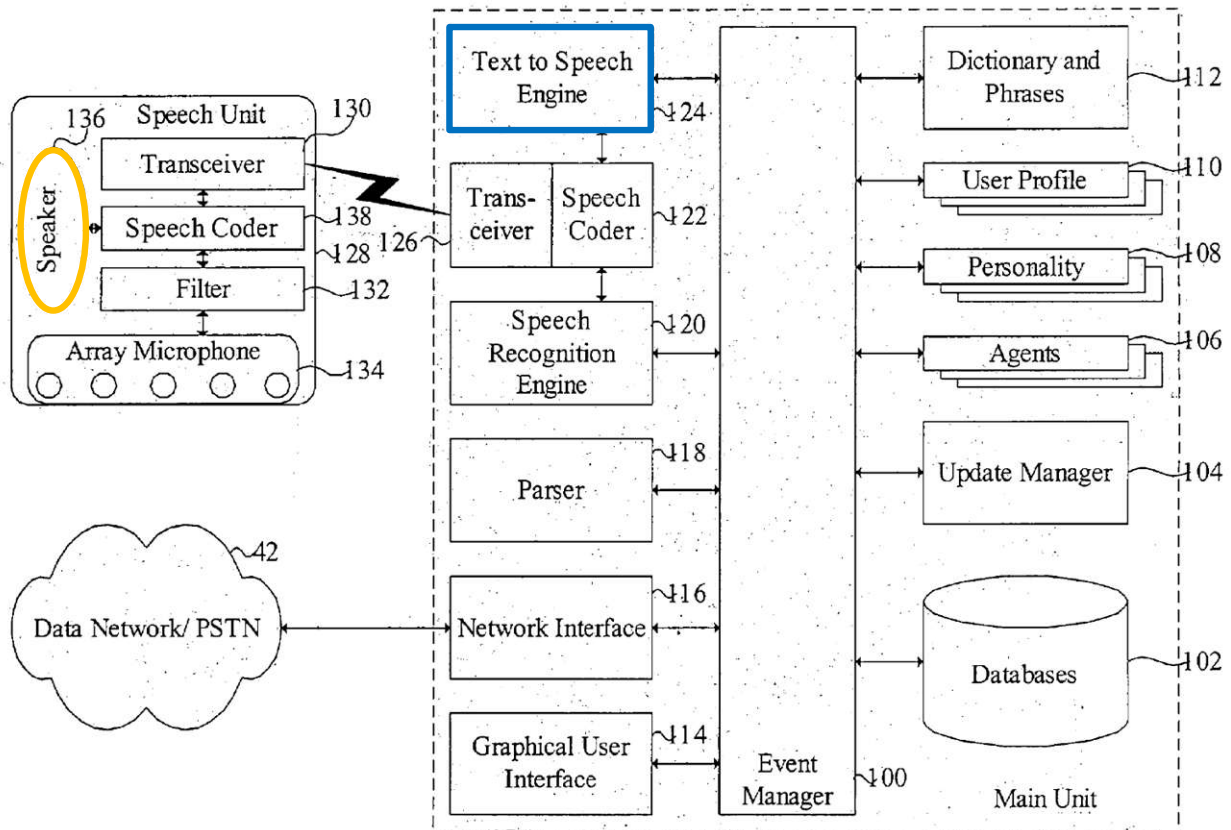


Figure 5. Speech Processing System Block Diagram

98

EX1004, Figure 5.

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

185. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 (“systems and methods are provided”). As I discuss for claim 3, Kennewick further discloses wherein “*the response comprises a voice response.*” Further, as I discuss for claim 3, Kennewick in combination with Cooper discloses “*wherein generating the voice response to the natural language utterance based on the*

identified manner comprises varying one or more of word use and/or jargon of the voice response.”

10. Claim 10

[10.pre] “The method of claim 1, further comprising:”

186. Claim 1 discloses a method as I discuss for [1.pre]. *See, e.g.*, EX1004, ¶11 (“systems and methods are provided”).

[10.a] “adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.”

187. As I discuss for limitation [1.h], Kennewick in combination with Cooper discloses generating “*the response*” to the natural language utterance.

188. Kennewick further discloses “*adapting the response to have a personality by varying word use*” by selecting a personality and transforming (adapting) responses to have that personality. EX1004, ¶¶24 (“a response that may be upsetting to the user is best presented in a sympathetic manner. In another example, information requiring immediate action or announcing a safety problem can be delivered with a definite and authoritative personality.”), 35, 141 (describing using personalities for responses by combining traits including “sarcasm, humor, irritation, and sympathy, and other traits”), 178-180. Kennewick’s adaptations include “*varying word use*” by performing “variable substitutions and transformations” to the response and “vary[ing]... the exact terminology used” in the response, “to create a more natural response to the user.” *Id.*, ¶178; *see also id.*,

¶180 (describing substitutions and transformations of a response to include “preferred terminology” based on the presentation personality). A POSITA would have understood that Kennewick’s computer system selecting a personality and transforming responses to have a selected personality by including preferred terminology is “*adapting the response to have a personality by varying word use.*”

189. Cooper also discloses “*adapting the response to have a personality by varying word use.*” As I discuss for limitation [6.a], Cooper varies word use to have a personality for its responses by (1) including words or phrases (response format) associated with polite discourse when politeness (identified manner) is detected in the speech input received from the user; and/or (2) including words or phrases associated with submissive discourse when anger is detected in the speech input received from the user. EX1005, 43:48-44:6. A POSITA would have understood that Cooper’s response adaptation by including certain words or phrases based on the identified manner is “*adapting the response to have a personality by varying word use.*”

190. Cooper further discloses its virtual assistant adapts the response “*to model a conversation.*” EX1005, 43:48-44:6. A POSITA would have understood that Cooper’s virtual assistant varying words in the output response based on words received in the input utterance is adapting the response “*to model a conversation*” because Cooper’s virtual assistant varies the words to imitate the user’s manner of

speaking. This understanding is consistent with the disclosure in the '699 patent, which describes modeling a conversation by showing an example conversation demonstrating how a response may imitate a user's input word choices. EX1001, 18:57-58, 19:1-19.

191. A POSITA would have found it obvious to incorporate Cooper's teachings regarding "*varying word use*" in the output response based on words received in the input utterance into Kennewick's system, to enable Kennewick's system to better "*model a conversation*" as taught by Cooper. For example, Kennewick discloses adapting responses to exhibit "personality characteristics," including using "preferred terminology" associated with different personalities. EX1004, ¶¶24, ¶180 (describing formatting a response to use "preferred terminology from the user profile or presentation personality 108.>"). Cooper further discloses modeling a conversation by varying the word use, so that the responses exhibit specific personality traits (e.g., politeness, submissiveness) that match or complement the manner of the user's inputs. *See* EX1005, 43:48-44:6. In the combination where Kennewick is modified to generate responses based on the identified manner of the user's utterance based on the teachings of Cooper, a POSITA would have found it obvious that it would adapt its responses to have a personality by varying word use based on Cooper's teachings of using words indicative of different personalities and types of discourse. In doing so, the system

would “model a conversation” by imitating and/or complementing the user’s manner of speaking as taught by Cooper, which enables the system to better “simulate the behavioral characteristics of real humans” in a conversation, and provide “responses to [the] user’s questions and commands seem more natural.” EX1004, ¶¶24, 35. Accordingly, Kennewick in combination with Cooper discloses “*adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.*”

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

192. Claim 1 discloses a method as I discuss for [1.pre]. See, e.g., EX1004, ¶11 (“systems and methods are provided”). Kennewick further discloses and renders obvious claim 11.

193. As I discuss for limitations [1.c], [1.d], and [1.h], Kennewick discloses generating a response to the natural language utterance based on the interpretation, which is based on the identified context. A POSITA would have understood that Kennewick generating a response based on the identified context amounts to “*generating a response that is sensitive to context.*” This is because a response that may vary based on context is a response that is sensitive to context.

194. Further, as I discuss for limitations [1.e], [1.f], and [1.h], Kennewick discloses generating a response based on (“*sensitive to*”) the “*short-term knowledge and long-term knowledge*.” Kennewick further discloses that its user profile contains long-term knowledge including user preferences, history of questions asked, and session histories. EX1004, ¶124 (“[t]he response can be created using the results of information queries, the system personality 108 and the user preferences or other data in the user profile 110.”); *see also id.*, ¶31 (“The system may use profile data in interpreting questions, formulating queries, interpreting results of queries and presenting answers to the user. Examples of information in a user profile includes, history of questions asked, session histories, formatting and presentation preferences, vehicle type, special vehicle equipment, vehicle related data, special word spelling, terms of interest, special data sources of interest, age, sex, education level, location of vehicle, planned path or route, specific addresses, commonly visited destinations, place of business, type of business, investments, hobbies, sports interests, news interests and other profile data.”).

195. A POSITA would have understood that Kennewick’s computer system generating a response based on the tagged dialog history (short-term knowledge) and the user profile (long-term knowledge) and in particular, based on “user preferences”, “presentation preferences”, and “history of questions asked” is

“generating a response that is sensitive to ... short-term knowledge and long-term knowledge of user preferences.”

196. Further, Kennewick’s use of the plural form in “session histories,” “history of their interactions with the system,” and “history of questions asked” confirms that the user profile in Kennewick stores information from multiple past conversations with the user. EX1004, ¶¶31, 143. A POSITA would have understood that Kennewick’s computer system generating a response based on this historical information shows that the generated response ***“is sensitive to ... words uttered by the user in one or more prior natural language utterances.”***

197. Kennewick further discloses that the response generated based on the system’s interpretation of the user’s natural language utterance is ***“sensitive to...what the user already knows about a topic.”*** For example, as part of interpreting the user’s utterance, Kennewick discloses determining “values for the criteria or parameters” that “have been explicitly supplied by the user” as part of a current or prior natural language utterance. EX1004, ¶162 (describing extracting “criteria or parameters from the user’s question or command” or from “the history of the dialog”). In addition, as discussed above, the generated response is “sensitive to” the contents of the user profile, which, in addition to long-term knowledge “added and updated automatically as the user uses the system,” contains information “manually added or updated by the user.” EX1004, ¶31. A POSITA would have

understood that information that the user “explicitly supplie[s]” as part of utterances in the conversation, or “manually add[s]” to the user’s profile, is information that is known to the user, and thus reflects *“what the user already knows about a topic.”* A POSITA would have also understood that Kennewick’s computer system generating a response based on this information provided in the conversation and/or in the user profile is *“generating a response that is sensitive to ... what the user already knows about a topic.”* See also EX1001, Abstract, 5:8-20 (describing short-term knowledge such as the dialog history and long-term knowledge stored in a user profile as “shared knowledge,” that is, knowledge known to the system and the user”).

198. Kennewick thus discloses generating a response based on (*“sensitive to”*) various pieces of information, including *“context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances,”* to “produce a response [that is] best understood and most natural to the user.” EX1004, ¶178. To produce such a response, Kennewick teaches making “maximum use of context, prior information,... and user specific profile data,” which includes the claimed *“context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances,”* to interpret the user’s

utterance and generate the response. EX1004, ¶118. It would have been obvious to a POSITA that to make “maximum use” of this information to generate a response that is best understood and most natural to the user, Kennewick’s system would consider all of this information when generating the response. Thus, the generated response will be based on (“*sensitive to*”) each of these pieces of information that is recited in the claim. Accordingly, a POSITA would have understood that Kennewick discloses and renders obvious “*generating a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.*”

B. Claims 12-22 are Rendered Obvious by Kennewick, Cooper, and Matsuda

199. Claims 12-22 correspond to claims 1-11 and are unpatentable for the same reasons. I provide additional explanation for why claims 12-22 are unpatentable below.

1. Claim 12

[12.pre] “A system for generating natural language system responses adapted based on a user’s manner of speaking, the system comprising:”

[12.a] “one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to:”

200. Claim 12 is unpatentable for the same reasons as claim 1. Preamble [12.pre] and limitation [12.a] are substantively identical to preamble [1.pre].

Preamble [12.pre] and limitation [12.a] recite “[a] system...comprising...one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to” perform recited operations. Preamble [1.pre] similarly recites “[a] computer-implemented method...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” comprising recited steps. The “computer system” implementing the method of claim 1 is equivalent to the “system” of claim 12 and obvious for the same reasons. The remaining limitations of claim 12 are substantively identical to the corresponding limitations of claim 1 and obvious for the same reasons.

2. Claims 13-22

201. Each of dependent claims 13-22 is unpatentable for the same reasons as the respective corresponding claim indicated in the following table:

Claim Number	Corresponding Claim Number
13	2
14	3
15	4
16	5
17	6

Claim Number	Corresponding Claim Number
18	7
19	8
20	9
21	10
22	11

202. Each of claims 13-22 is substantively identical to the corresponding claim and contains only minor differences to reflect that independent claim 12 recites a system. Accordingly, claims 13-22 are unpatentable for the same reasons I provide for claim 12 and the respective corresponding claim.

X. CONCLUSION

203. For the reasons I provide above, it is my opinion that claims 1-22 of the '699 patent are unpatentable.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: 8/4/2025

Padhraic Smyth
Padhraic Smyth