

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

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 10,755,699**

TABLE OF CONTENTS

	Page
I. OVERVIEW OF THE '699 PATENT	1
A. Alleged Invention	1
B. Priority Date	3
II. IDENTIFICATION OF CHALLENGE	3
A. Statutory Grounds.....	3
B. Prior Art.....	4
1. Kennewick.....	4
2. Cooper	6
3. Matsuda	7
III. LEVEL OF ORDINARY SKILL IN THE ART	8
IV. CLAIM CONSTRUCTION	8
V. GROUND 1: Kennewick, Cooper, and Matsuda	9
A. Claim 1	9
[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:”	9
[1.a]: “receiving, by the computer system, a user input comprising a natural language utterance;”	15

TABLE OF CONTENTS
(Continued)

	Page
[1.b] “recognizing, by the computer system, one or more words or phrases from the natural language utterance;”	16
[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;”	18
[1.d] “determining, by the computer system, an interpretation of the natural language utterance based on the identified context;”	20
[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;”	21
[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;”	26
[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”	32
[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.”	37

TABLE OF CONTENTS (Continued)

	Page
B. 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.”	39
C. 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.”	40
D. Claim 4	43
[4.pre] “The method of claim 1, the method further comprising:”	43
[4.a] “obtaining, by the computer system, contextual signifiers and/or grammatical rules,”	43
[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.”	44
E. Claim 5	46
[5.pre] “The method of claim 1, wherein the long-term knowledge is associated with a first user, the method further comprising:”	46
[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.”	46

TABLE OF CONTENTS
(Continued)

	Page
F. Claim 6: “The method of claim 1, wherein generating the response based on the identified manner comprises: adapting, by the computer system, the response based on a response format associated with the identified manner.”.....	47
G. Claim 7	49
[7.pre] “The method of claim 1, wherein recognizing the one or more words or phrases from the natural language utterance comprises:”	49
[7.a] “providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and”	49
[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.”	49
H. Claim 8: “The method of claim 1, the method further comprising: causing, by the computer system, the response to the natural language utterance to be provided to the user.”	50
I. 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.”	51
J. Claim 10: “The method of claim 1, further comprising: adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.”	52

TABLE OF CONTENTS
(Continued)

	Page
K. 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.”	54
L. Claim 12	56
[12.pre] “A system for generating natural language system responses adapted based on a user's manner of speaking, the system comprising:”	56
[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:”	56
M. Claims 13-22	57
VI. CONCLUSION.....	58
VII. STANDING.....	58
VIII. MANDATORY NOTICES AND FEES	58
A. Real Party-in-Interest	58
B. Related Matters.....	58
C. Lead and Backup Counsel.....	60
D. Service Information	60
E. Fees.....	61

TABLE OF AUTHORITIES

	Page(s)
STATUTES AND RULES	
35 U.S.C. § 102	4, 6, 7
35 U.S.C. § 103	4

EXHIBIT LIST (37 C.F.R. § 42.63(e))

Exhibit	Description
1001	U.S. Patent No. 10,755,699 (“the ’699 patent”)
1002	Declaration of Padhraic Smyth, Ph.D.
1003	Prosecution File History for the ’699 patent
1004	U.S. Patent App. Pub. No. 2004/0193420 (“Kennewick”)
1005	U.S. Patent No. 6,757,362 (“Cooper”)
1006	Japanese Patent App. Pub. No. 2004/334591 (“Matsuda”)
1007	Certified English Translation of Japanese Patent App. Pub. No. 2004/334591 (“Matsuda”)
1008	Excerpt of The American Heritage Dictionary of the English Language, 4th Ed. (2000), Houghton Mifflin Co., ISBN 0-395-82517-2
1009	Asthana, A. & Krzyzanowski, P., “A Small Domain Communications System for Personalized Shopping Assistance,” Proceedings of 1994 International Conference on Personal Wireless Communications, IEEE Press 199-203 (Aug. 1994) (“Asthana”)
1010	Lucente, M., “Conversational interfaces for e-commerce applications,” Communications of the ACM 43(9) 59-61 (Sept. 2000) (“Lucente”)
1011	Huang, X. et al., “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development,” Prentice Hall (2001) (“Huang”)
1012	U.S. Patent No. 6,859,776 (“Cohen”)
1013	Complaint for Patent Infringement, <i>VB Assets, LLC, v. Amazon.com Services, LLC</i> , 1:24-cv-00839-MN (D. Del. July 18, 2024), Dkt. 1
1014	Curriculum Vitae of Padhraic Smyth, Ph.D.
1015	Claim Construction Memorandum Order, <i>VB Assets, LLC, v. Amazon.com, Inc. et al.</i> , 1:19-cv-01410-MN (D. Del. June 23, 2021), Dkt. 90

CHART OF CLAIMS
[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:
[1.a] receiving, by the computer system, a user input comprising a natural language utterance;
[1.b] recognizing, by the computer system, one or more words or phrases from the natural language utterance;
[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;
[1.d] 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;
[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;
[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
[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.
[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.

CHART OF CLAIMS
[3] The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response based on the identified manner comprises varying one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.
[4.pre] The method of claim 1, the method further comprising:
[4.a] obtaining, by the computer system, contextual signifiers and/or grammatical rules,
[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.
[5.pre] The method of claim 1, wherein the long-term knowledge is associated with a first user, the method further comprising:
[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.
[6.pre] The method of claim 1, wherein generating the response based on the identified manner comprises:
[6.a] adapting, by the computer system, the response based on a response format associated with the identified manner.
[7.pre] The method of claim 1, wherein recognizing the one or more words or phrases from the natural language utterance comprises:
[7.a] providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and
[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.

CHART OF CLAIMS	
[8.pre]	The method of claim 1, the method further comprising:
[8.a]	causing, by the computer system, the response to the natural language utterance to be provided to the user.
[9]	The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response to the natural language utterance based on the identified manner comprises varying one or more of word use and/or jargon of the voice response.
[10.pre]	The method of claim 1, the method further comprising:
[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.
[11]	The method of claim 1, the method further comprising generating a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.
[12.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:
[12.a(i)]	receive a user input comprising a natural language utterance;
[12.a(ii)]	recognize one or more words or phrases from the natural language utterance;
[12.a(iii)]	identify a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;
[12.a(iv)]	determine an interpretation of the natural language utterance based on the identified context;

CHART OF CLAIMS
[12.a(v)] accumulate short-term knowledge based on one or more natural language utterances received during a predetermined time period, wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;
[12.a(vi)] accumulate long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;
[12.a(vii)] identify a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and
[12.a(viii)] generate a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.
[13] The system of claim 12, wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.
[14.pre] The system of claim 12, wherein the response comprises a voice response, and wherein to generate the voice response based on the identified manner, the one or more physical processors are further configured to:
[14.a] vary one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.
[15.pre] The system of claim 12, wherein the one or more physical processors are further configured to:
[15.a] obtain contextual signifiers and/or grammatical rules,
[15.b] wherein to generate the response based on the identified manner in which the natural language utterance was spoken, the one or more physical processors are further configured to use the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.

CHART OF CLAIMS
[16.pre] The system of claim 12, wherein the long-term knowledge is associated with a first user, and wherein the one or more physical processors are further configured to:
[16.a] generate a profile associated with the first user based on the long-term knowledge, wherein the context for the natural language utterance is determined based further on the profile associated with the first user.
[17.pre] The system of claim 12, wherein to generate the response based on the identified manner, the one or more physical processors are configured to:
[17.a] adapt the response based on a response format associated with the identified manner.
[18.pre] The system of claim 12, wherein to recognize the one or more words or phrases from the natural language utterance, the one or more physical processors are configured to:
[18.a] provide the natural language utterance as an input to a speech recognition engine; and
[18.b] obtain the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.
[19.pre] The system of claim 12, wherein the one or more physical processors are further configured to:
[19.a] cause the response to the natural language utterance to be provided to the user.
[20.pre] The system of claim 12, wherein the response comprises a voice response, and wherein to generate the voice response to the natural language utterance based on the identified manner, the one or more physical processors are further configured to:
[20.a] vary one or more of word use and/or jargon of the voice response.
[21.pre] The system of claim 12, wherein the one or more physical processors are further configured to:

CHART OF CLAIMS
[21.a] adapt the response to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.
[22] The system of claim 12, wherein the one or more physical processors are further configured to generate a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.

U.S. Patent No. 10,755,699 (the “’699 patent”) claims technology for facilitating natural language interactions with a computer system, where the computer system generates responses “adapted based on a user’s manner of speaking.” EX1001, 20:35-37. The claims also recite the responses being based on accumulated knowledge. *Id.*, 20:53-63. But this technology was well-known in the prior art. *See, e.g.*, EX1002, ¶¶50-71. This Petition demonstrates that the prior art renders each claim of the ’699 patent obvious. *See also* EX1002, ¶¶1-203.

I. OVERVIEW OF THE ’699 PATENT

A. Alleged Invention

The ’699 patent is purportedly directed to an interface that “understand[s] free form human utterances” spoken by a user. EX1001, 2:17-21, 11:30-35; EX1002, ¶¶32-36. The interface uses “context” to disambiguate utterances and improve interpretation accuracy. EX1001, 4:32-36; EX1002, ¶33. For example, the word “traffic” may refer to a road condition, a music band, or a movie, depending on the context. EX1001, 11:45-55; EX1002, ¶33. The context may be used within a “cooperative conversation model...to implement a conversational Human-to-Machine interface that reflects how humans interact with each other and their normal behavior in conversation.” EX1001, 13:51-56; EX1002, ¶33.

The ’699 patent describes an Intelligent Hypothesis Builder that “generate[s] one or more explicit hypotheses of a user’s intent.” EX1001, 16:43-47; EX1002,

¶34. 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.” EX1001, 16:24-31; EX1002, ¶34. “Short-term knowledge may accumulate during a single conversation, where input received during a single conversation may be retained.” EX1001, 5:10-12; EX1002, ¶34. 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.” EX1001, 5:29-33; EX1002, ¶34.

The system may use both short-term and long-term knowledge about the user to generate a response. EX1001, 14:3-5; EX1002, ¶35. 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...that facilitate interaction at a Human-to-Machine interface.” EX1001, 14:64-15:3; EX1002, ¶35. An Adaptive Response Builder uses the output of the Intelligent Hypothesis Builder “to formulate [a] personalized and conversational response” to the user, which “may include a verbal or audible reply played through an output device..., and/or an action performed by a device, computer, or machine.” EX1001, 16:63-66, 17:41-46, 18:9-14; EX1002, ¶35. The response may be further based on the “user’s manner of speaking” and/or formatted to exhibit “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).” EX1001, 18:45-50; EX1002, ¶36.

B. Priority Date

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 filed on April 20, 2015, now Patent No. 10,297,249, which is a continuation of Application No. 13,987,645, filed on August 19, 2013, now Patent No. 9,015,049, which is a division of Application No. 13/251,712 (the “’712 application”) filed on October 2, 2011, now Patent No. 8,515,765, which is a continuation-in-part of Application No. 11/580,926 (the “’926 application”) filed on October 16, 2006 (the effective filing date), now Patent No. 8,073,681. EX1001; *see also* EX1002, ¶31. This Petition shows the challenged claims were invalid as of the effective filing date.¹

II. IDENTIFICATION OF CHALLENGE

A. Statutory Grounds

Petitioner requests *inter partes* review and cancellation of the challenged claims on the following ground. *See also* EX1002, ¶¶1-203.

¹ Petitioner reserves the right to challenge the ’699 patent’s priority date in this or other proceedings.

Ground	Claims	Statutory Basis	Prior Art
1	Claims 1-22	§ 103	Kennewick, Cooper, Matsuda

B. Prior Art²

1. Kennewick

Kennewick is a U.S. patent application published September 30, 2004. EX1004. Kennewick is prior art under at least pre-AIA 35 U.S.C. § 102(b) because it published more than one year before the '699 patent's priority date. EX1004; EX1002, ¶61.

Kennewick describes a system that allows “users to ask natural language speech questions” and “present[s] results in a natural manner.” EX1004, ¶2; EX1002, ¶62. The 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.” EX1002, ¶11; EX1002, ¶62. The system receives input using a speech unit, “processes the input..., and responds to the user with a natural language speech response.” EX1004, ¶¶12, 14; EX1002, ¶62.

² Pre-AIA 35 U.S.C. § 102 applies to the '699 patent.

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’).” EX1004, ¶156; EX1002, ¶63. Kennewick discloses tagging words and phrases recognized for each utterance of the dialog, to “allow other components of the system to tie that utterance to the correct...dialog.” EX1004, ¶155; EX1002, ¶63. The tagged information forms a “dialog history,” which Kennewick’s system uses to interpret the utterance and generate an appropriate response. EX1004, ¶¶157, 160, 162, 164, 170; EX1002, ¶63.

Kennewick’s system further “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.” EX1004, Abstract, ¶31; *see also id.*, ¶160 (describing “determin[ing] a context for an utterance” based on a “user profile,” among other values); EX1002, ¶64. The “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.” EX1004, ¶¶31; EX1002, ¶64. 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 appropriate responses. EX1004, ¶¶157, 160, 162, 164, 170; EX1002, ¶64. Kennewick’s system further formats the responses to “simulate

some aspects of human ‘personality,’” so that responses “do not appear rigidly formatted or mechanical.” EX1004, ¶24; EX1002, ¶64.

2. *Cooper*

Cooper is a U.S. patent issued June 29, 2004. EX1015. Cooper is prior art under at least pre-AIA 35 U.S.C. § 102(a) and (b) because its issue date predates the priority date of the ’699 patent. EX1005; EX1002, ¶65.

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; EX1002, ¶66. Cooper determines the user’s manner of speech and adapts the assistant’s output to match the user. EX1005, 43:51-61; EX1002, ¶¶66-67. For example, if the user uses polite words when speaking to the assistant, the assistant may use polite words when responding to the user. *Id.* In addition, the virtual assistant may use “information about the user’s voice volume, word choice and speech rate” included in the user’s input to determine “the user’s emotional state, calm or angry, for example.” EX1005, 43:62-44:4; EX1002, ¶67. If “the user’s emotional state is angry,” the virtual assistant may automatically generate responses that “include words associated with submissive discourse, such as ‘sorry,’ ‘regret’ and ‘apologize.’” *Id.* Moreover, the “virtual assistant could save the information obtained about the user’s emotional state for use in future sessions.” EX1005, 44:4-6; EX1002, ¶67.

3. Matsuda

Matsuda is a Japanese patent application published November 25, 2004. EX1006. Exhibit 1007 is a certified English translation of Matsuda. Matsuda is prior art under at least pre-AIA 35 U.S.C. § 102(a) and (b) because its publication date predates the priority date of the '699 patent. EX1006; EX1007; EX1002, ¶68.

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. EX1007, ¶26; EX1002, ¶69. 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,” in which case “the CPU 11 sets the ‘subject person number’ to 0 [] and clears the ‘subject person number latest setting date/time.’” EX1007, ¶¶26, 38; EX1002, ¶69. Matsuda explains that this technique enables the system to better recognize subjects referred to by the user earlier in the conversation, even when the user does not use exactly the same name. EX1007, ¶¶80, 97; EX1002, ¶69.

Matsuda further discloses retaining name-related information across conversations in a persistent data structure. EX1007, ¶¶31, 86; EX1002, ¶70. When the user utters a subject name, the system compares the spoken name against existing entries in the data structure and if a match is found, sets the person number of the matching entry as the current subject. EX1007, ¶76; EX1002, ¶70. If no match is

found, the system determines that the named person has not been mentioned. EX1007, ¶74; EX1002, ¶70. 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* EX1007, ¶¶77-82, EX1002, ¶70. 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; EX1002, ¶¶59, 71.

III. LEVEL OF ORDINARY SKILL IN THE ART

A POSITA at the priority date of the '699 patent would have had a Bachelor's 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. Additional experience can substitute for the level of education, and vice-versa. EX1002, ¶¶3-13, 23-26.

IV. CLAIM CONSTRUCTION

For the purposes of this Petition, Petitioner applies plain-and-ordinary meaning to all terms.³ *See also* EX1002, ¶¶27-30.

³ Petitioner reserves the right to offer different claim constructions in other forums.

V. GROUND 1: KENNEWICK, COOPER, AND MATSUDA

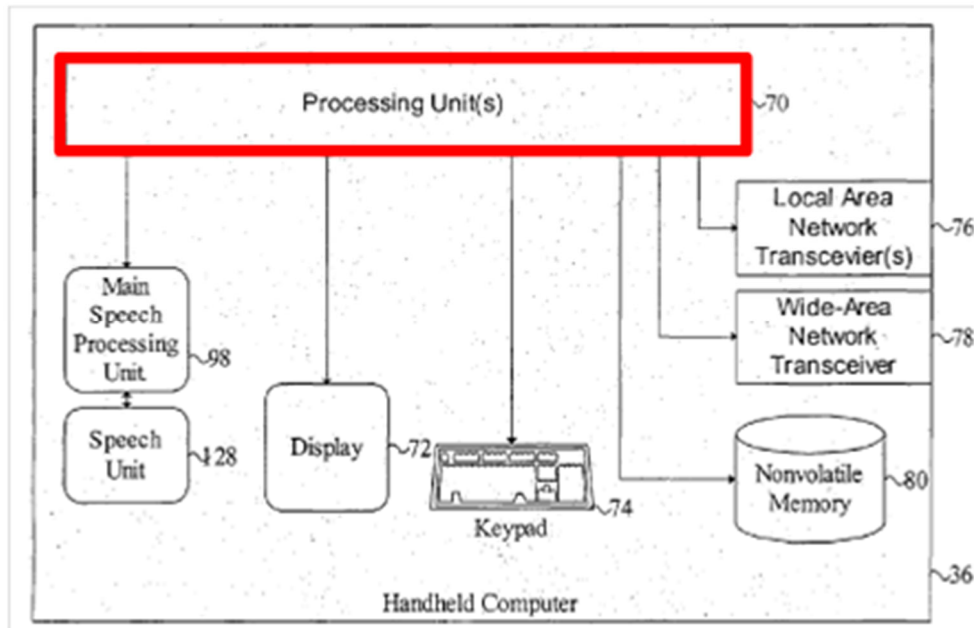
A. Claim 1

[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:”

Kennewick in combination with Cooper discloses the preamble. EX1002, ¶¶74-90. Kennewick describes 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; EX1002, ¶74. The environment “responds to the user with a natural language speech response.” EX1004, ¶¶14, 23, 24, 27, 28, 118, 155-56, claim 28; EX1002, ¶¶74-75. The environment is provided by “systems and methods” and therefore performs a “*method of generating natural language system responses*.” EX1004, ¶11; EX1002, ¶75.

Kennewick also discloses that the method is “*computer-implemented...by a computer system*.” EX1002, ¶¶76-77. Kennewick’s disclosed method may be implemented on numerous “computer alternatives” that Kennewick collectively refers to as “a computer.” EX1004, ¶13; EX1002, ¶76. The computer may be a

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



EX1004, Figure 3, ¶¶90; EX1002, ¶76. Kennewick’s Figure 4 is a near-identical figure illustrating an embodiment of fixed computer. EX1004, Figure 4, ¶¶91; EX1002, ¶76. These embodiments of the computer are each a “*computer system*” because they are formed of various computing components working together. *See* EX1001, ¶¶3, 96, 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”); EX1002, ¶77.

Kennewick’s “*computer system...includes one or more physical processors.*” EX1002, ¶78. Figure 3, shown above, illustrates that the computer system has “*processing units,*” which “may include one or more central processing

units, one or more data and address busses, data interfaces and volatile memory.”

EX1004, ¶¶111, 115, 101, 122, 107-09; EX1002, ¶78. Kennewick’s computer system thus “*includes one or more physical processors.*” EX1002, ¶78.

The processors “*execut[e] one or more computer program instructions which, when executed, perform the method.*” EX1002, ¶79. Kennewick discloses that the computer system executes software to perform the method, including “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.” EX1004, ¶13; EX1002, ¶79. The software is formed of computer program instructions which the physical processors execute to perform the method. *See, e.g.*, EX1004, ¶¶17, 107; EX1002, ¶79.

Kennewick therefore discloses “[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.*” EX1002, ¶80.

Kennewick also discloses using different personalities for responses, allowing the system to exhibit personality traits such as “sarcasm, humor, irritation, and sympathy, and other traits.” EX1004, ¶141; EX1002, ¶81. Kennewick states that this “use of simulated personality characteristics is...desirable,” and gives examples

where the system may select different personalities, such as presenting upsetting information “in a sympathetic manner” or urgent alerts “with a definite and authoritative personality.” EX1004, ¶24, ¶141; EX1002, ¶81. The system may select different personality characteristics for the response based on a variety of factors such as the “context” of the conversation and information about the user stored in a user profile. EX1004, ¶179; EX1002, ¶81. Further, Kennewick’s system tailors and formats its responses based on domain knowledge and user preferences. EX1004, ¶¶17, 23, 24; EX1002, ¶82.

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, which describes a “virtual assistant [that] adapts to the user in many different ways based on the [received] input.” EX1005, 2:26-30, 2:56-64, Abstract; EX1002, ¶¶83-84. For example, Cooper’s virtual assistant can consider “whether words associated with polite discourse are included in input received from the user,” and if so, “automatically enable [a] politeness setting” to “cause the virtual assistant to include words or phrases associated with polite discourse in the output from the virtual assistant.” EX1005, 43:49-61; EX1002, ¶84. In addition, the virtual assistant may receive “information about the user’s emotion, which could be based on information about the user’s voice volume, word choice and speech rate.” EX1005, 43:62-67; EX1002, ¶85. The system may, “[b]ased on such information,...

automatically determine the user's emotional state, calm or angry, for example" and generate a matching response. *Id.* For example, "[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.'" EX1005, 44:1-4; EX1002, ¶¶84-85. Cooper thus 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. EX1005, Abstract; EX1002, ¶86. This 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" and "create a conversational feel." EX1001, 6:32-34, 18:45-50; EX1002, ¶86.

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***." EX1002, ¶¶87-90. Kennewick and Cooper are analogous art, both relating to voice-based interactions with computers. *See, e.g.*, EX1004, ¶2; EX1005, 2:37-39; EX1002, ¶87. Kennewick discloses that its 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; EX1002, ¶88.

Kennewick system further tailors its responses based on domain knowledge and user preferences. EX1004, ¶¶17, 23, 24; EX1002, ¶88. Likewise, Cooper teaches that its virtual assistant automatically “adapts to the user in many different ways based on the input the virtual assistant receives,” by adjusting the manner of its responses based on the manner of the user’s input, e.g., enabling the politeness setting if the user uses words of polite discourse, or “automatically includ[ing] words associated with submissive discourse” if “the user’s emotional state is angry.” EX1005, 2:56-63, 43:49-61; 44:1-4; EX1002, ¶88.

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 system that uses different personalities for responses and tailors responses based on domain knowledge and user preferences. EX1002, ¶89. 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, consistent with Kennewick's stated objective of making "responses to the user's questions and commands seem more natural." EX1004, ¶¶24, 35; EX1002, ¶89.

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. EX1002, ¶90. The combination is also a predictable use of known techniques for improving voice-based interactions with computers. EX1002, ¶90. The improvement would have been within the skill of a POSITA, who would have reasonably expected success based on the compatibility of the systems and their shared goal of enabling adaptive, conversational interfaces. EX1002, ¶90.

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

Kennewick discloses limitation [1.a]. EX1002, ¶¶91-92. Kennewick's "*computer system*" includes a **speech unit**. EX1004, ¶12; EX1002, ¶91.

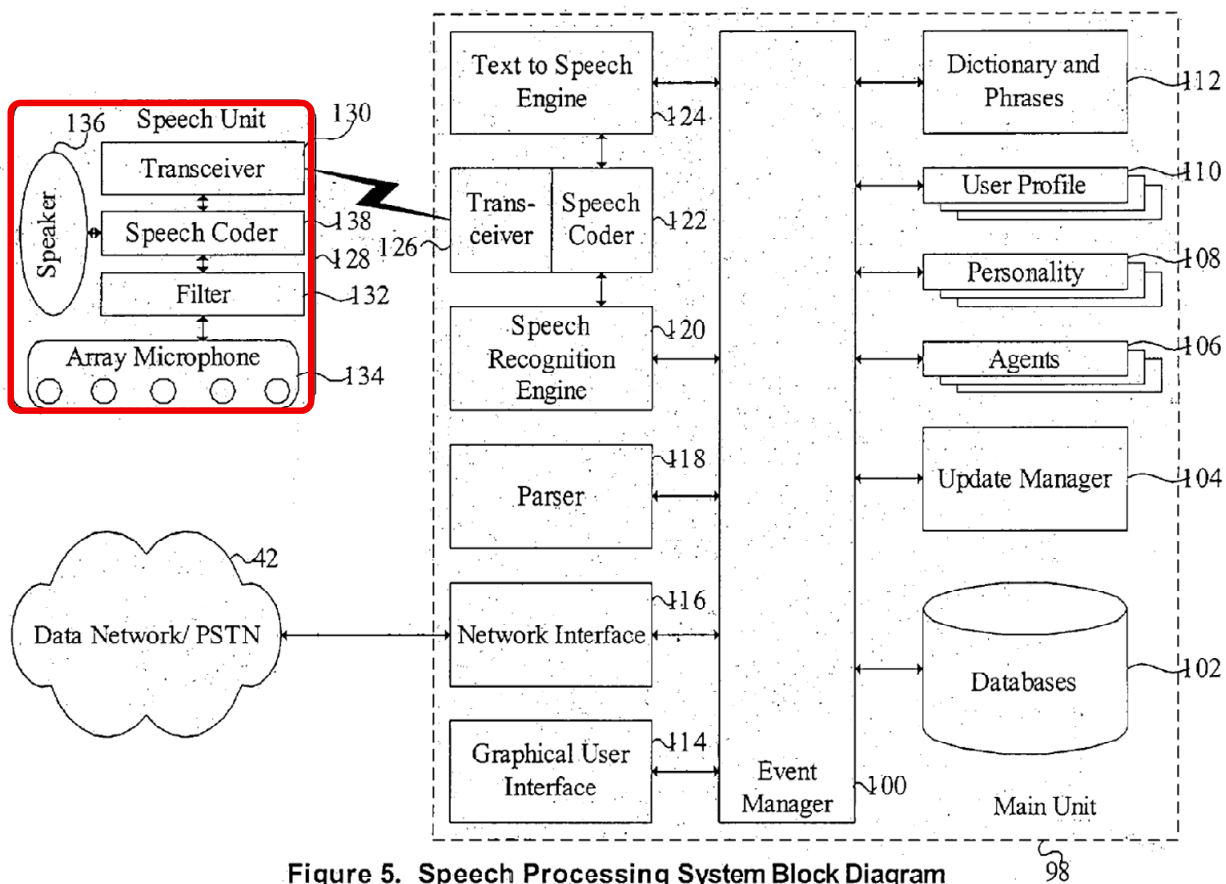


Figure 5. Speech Processing System Block Diagram

EX1004, Fig. 5. The “**speech unit 128** includes one or more microphones...to receive the utterances of the user,” which are “*a user input comprising a natural language utterance.*” EX1004, ¶¶14 (“spoken natural language queries, commands and/or other utterances”), 121; EX1002, ¶92.

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

Kennewick discloses limitation [1.b]. EX1002, ¶¶93-95. The received natural language utterance is coded by the speech coder 138 in the **speech unit 128** and transmitted to the main unit 98. EX1004, ¶121; EX1002, ¶93. From there, Kennewick discloses passing the coded speech “to the speech coder 122 [in the main

unit 98] for decoding and decompression,” and feeding the output “to the **speech recognition engine 120.**” EX1004, ¶¶123, 155; EX1002, ¶¶93-94.

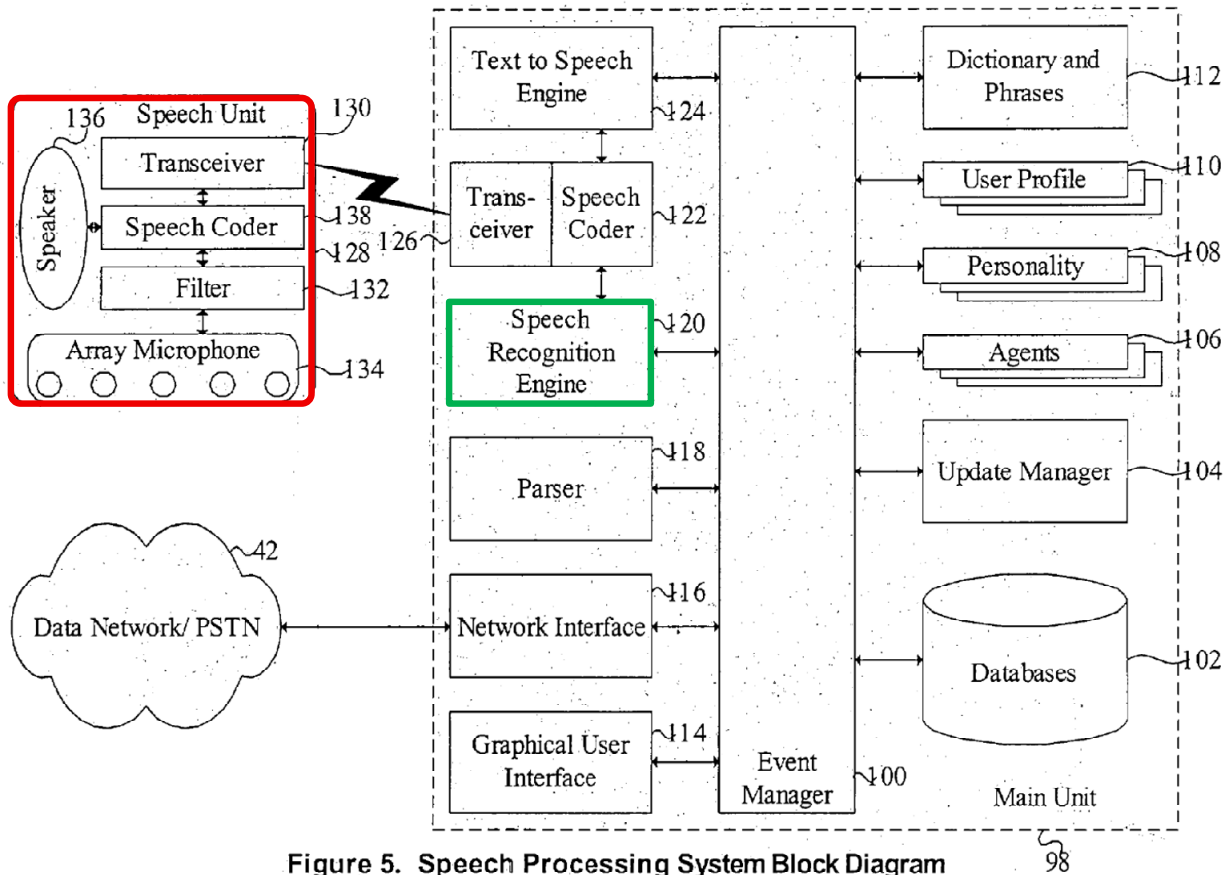


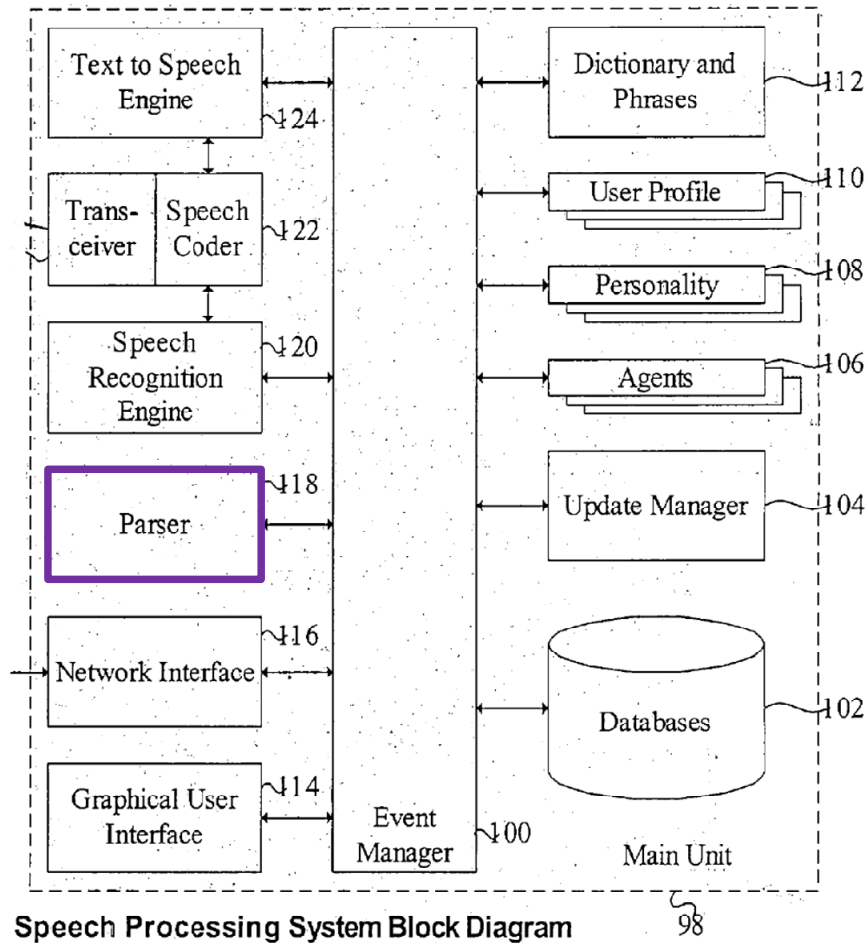
Figure 5. Speech Processing System Block Diagram

The “**speech recognition engine 120**” recognizes “words and phrases, using for example, information in the dictionary and phrase tables 112, and pass[es] these to the parser 118 for interpretation.” EX1004, ¶123, 155, Fig. 5; EX1002, ¶94. Kennewick’s speech recognition engine’s 120 analysis of 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.*” EX1004, ¶¶ 121, 123, 155; EX1002, ¶95.

[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;”

Kennewick discloses limitation [1.c]. EX1002, ¶¶96-98. As discussed for limitations [1.a] and [1.b], Kennewick’s “***computer system***” receives a “***natural language utterance***,” and recognizes “***one or more words or phrases...from the natural language utterance.***” EX1004, ¶¶123, 155; EX1002, ¶96. Kennewick further teaches that the 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.***”

As shown in Figure 5, Kennewick’s speech processing unit includes a **parser 118** that determines (identifies) a context based on the words and phrases recognized by the speech recognition engine 120.



EX1004, ¶160, Figure 5 (excerpt); EX1002, ¶¶96-97. “Once the words and phrases have been recognized by the speech recognition engine 120, the tokens” are “passed to the parser 118,” which “examines the tokens for the questions or commands, context and criteria,” and “determine[s] a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching.” *Id.* Thus, Kennewick’s parser 118 identifies the “context” “based on” the recognized words or phrases from the user’s natural language utterance (i.e., “tokens”). *Id.* Accordingly, Kennewick discloses and renders 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.” EX1002, ¶98.

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

Kennewick discloses limitation [1.d]. EX1002, ¶¶99-102. Kennewick’s parser interprets natural language utterances received from the user, based on the contexts of the utterances. EX1004, ¶¶7, 28, 41, 143, 155; EX1002, ¶¶99-100. In particular, the parser uses recognized words and phrases, the identified context, and other criteria to determine the appropriate context-based agent to be invoked for a user utterance forming a question or command. The agent works with the parser to interpret the utterance based on the context. EX1004, ¶¶160 (“[t]he context of a question or command may determine the domain and thereby, the domain agent 156, if any, to be invoked”), 162-163, 165; EX1002, ¶¶100-101. 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.” EX1004, ¶163; EX1002, ¶101. Thus, because Kennewick describes its computer system determining the correct agent for interpreting the utterance 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.*” EX1002, ¶102.

[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;”

Kennewick in combination with Matsuda discloses limitation [1.e]. EX1002, ¶¶103-117. Kennewick discloses “*a single conversation between a user and the computer system.*” The user and computer system have a “dialog” starting with the user addressing the computer system and ending with the user uttering a dismissal word. EX1004, ¶156; EX1002, ¶104. The dialog between the addressal and dismissal words is a “*single conversation between [the] user and the computer system.*” EX1002, ¶104; *see also* EX1008 (dictionary defining “dialogue or dialog” as “A conversation between two or more people.”).

Kennewick also discloses that the “*computer system*” “*accumulat[es]...short-term knowledge based on one or more natural language utterances...related to [the] single conversation.*” EX1002, ¶¶105-108.

Kennewick’s system builds a dialog history that is “*short-term knowledge*” because it reflects details about a user’s utterances during the dialog. EX1004, ¶¶23, 25, 29, 37, 155, 157, 160; EX1002, ¶¶106-108; *see also* EX1001, 16:50-52 (describing “short-term knowledge” as including a “history of the conversation” which is

accumulated “as the conversation progresses”). Kennewick’s computer system “*accumulat[es]*” the short-term knowledge by tagging each natural language utterance received during the dialog, “to tie that utterance to the correct...dialog” and user. EX1004, ¶¶155-56; EX1002, ¶¶107-108. The knowledge is related to the natural language utterances received as user input during the dialog. EX1002, ¶¶107-108. Accordingly, Kennewick discloses “*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.*” EX1002, ¶108.

Kennewick does not explicitly disclose accumulating short-term knowledge based on one or more natural language utterances “*received during a predetermined time period.*” EX1002, ¶¶109-110. Matsuda does. Matsuda discloses a conversation system that receives natural language utterances from a user. EX1007, ¶36; EX1002, ¶110. The system stores information about a person who is the subject of a conversation in a memory, which enables the conversation “to progress naturally while returning appropriate responses in relation to the subject.” EX1007, ¶12; EX1002, ¶111.

Matsuda’s system “*accumulat[es]...short-term knowledge*” about the subject person “*based on one or more natural language utterances received during a*

predetermined time period.” It stores name data describing people who are or may be the subject of a conversation with the user. EX1007, ¶31; EX1002, ¶110. The name data includes a “person number” and one or more names for a person. *Id.* The system also stores a “subject person number” indicating the subject of a current conversation between the system and a user. EX1007, ¶¶23, 26, 39; EX1002, ¶111. If the subject person number is 0 there is no current subject person. EX1007, ¶26; EX1002, ¶111. Otherwise, the person having that person number is the subject. *Id.*

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; EX1002, ¶112. When the subject person number is established, the system tracks elapsed time since the subject person was last mentioned by the user. EX1007, ¶¶27, 69; EX1002, ¶112. The system expires the subject person number if the elapsed time exceeds “*a predetermined time period.*” EX1007, ¶¶27, 38, 69. Specifically, “if a prescribed time (for example, 10 minutes) has elapsed since” the subject person was last mentioned, Matsuda’s system sets the subject person number to 0 and clears the timer tracking the elapsed time. EX1007, ¶¶27, 38, 69; EX1002, ¶112. This way, the system avoids referencing the wrong person in the conversation. *See* EX1007, ¶¶95, 100. EX1002, ¶112.

During the “*predetermined time period*,” Matsuda’s system further “*accumulat[es]*” name information about the subject person derived from the “natural language utterances” it receives from the user during the conversation. EX1007, ¶¶62-67; EX1002, ¶¶113-114. If the user utters a name similar to the subject person’s name, Matsuda’s system treats the similar name as an additional name for that person. EX1007, ¶¶31, 62-66, Fig. 9; EX1002, ¶113. For example, if the conversation subject is “Matsui of the Kyo**” and user utters “Matsui is strong” during the “prescribed time” (“*predetermined time period*”), the system infers that “Matsui” is the same person as “Matsui of the Kyo**.” It therefore updates the stored information to indicate that “Matsui” is another name for “Matsui of the Kyo**” by using the same person number for both names. EX1007, ¶¶64, 74-75; EX1002, ¶113. This accumulated information is “*short-term knowledge*” because it is knowledge accumulated “during a single conversation” “as the conversation progresses.” EX1001, 5:10-12, 14:5-6, 16:59-62, 17:26-29; EX1002, ¶114. Matsuda’s system therefore “*accumulates[es]...short-term knowledge based on one or more natural language utterances received during a predetermined time period.*”

A POSITA would have been motivated to combine Kennewick and Matsuda, and the combination discloses “*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.” EX1002, ¶¶115-117. 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. EX1002, ¶115. Kennewick’s system builds a “user dialog history” which is short-term knowledge for a conversation. EX1004, ¶¶155-162; EX1002, ¶115. Matsuda’s system describes accumulating short-term knowledge for a conversation during a predetermined time period. EX1007, ¶¶26, 75; EX1002, ¶115.

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 the predetermined time period into Kennewick’s system. EX1002, ¶116. Kennewick’s system uses available information “to achieve a natural environment for one or more users making queries or commands in multiple domains.” EX1004, ¶11; EX1002, ¶116. 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; EX1002, ¶116. Accordingly, a POSITA would have been motivated to incorporate Matsuda’s accumulated

information into Kennewick's short-term knowledge in order 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; EX1002, ¶116.

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. EX1004, ¶118; EX1002, ¶117. 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. EX1002, ¶117.

[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;"

Kennewick in combination with Matsuda discloses limitation [1.f]. EX1002, ¶¶118-129. As discussed for limitation [1.e], Kennewick in combination with Matsuda discloses 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.” EX1002, ¶118.

First, Kennewick discloses the “*computer system*” “*accumulating...long-term knowledge*,” which the ’699 patent describes as “generally...user-centric, rather than session-based” knowledge that “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.” EX1001, 5:29-39; EX1002, ¶119. Kennewick’s “*computer system*” similarly “*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; EX1002, ¶120. 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.” EX1004, ¶143; EX1002, ¶120. The historical information in the profile is unconstrained by the current session, and “may be added and updated automatically as the user uses the system or can be manually added or updated by the user.” *See, e.g.*, EX1004, ¶31; EX1002, ¶120. 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.*” EX1002, ¶120.

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]. EX1002, ¶121. As discussed above, Kennewick discloses that conversations have beginnings and ends, and 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*”) the single conversation, where the computer system receives natural language utterances from the user during each conversation. See EX1004, ¶156; EX1002, ¶121. 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 the user’s “special use of terminology.” EX1004, ¶143; EX1002, ¶121.

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. See EX1002, ¶¶122-126. As discussed for limitation [1.e], Matsuda discloses tracking the elapsed time since the subject of a conversation was last mentioned by the user, and removing the subject

if the elapsed time exceeds the “*predetermined time period*.” EX1007, ¶¶26, 27, 38; EX1002, ¶123.

Matsuda’s system further “*accumulat[es]...long-term knowledge...based on one or more natural language utterances*.” Matsuda’s system stores name data (e.g., person names and numbers) in a nonvolatile memory that accumulates over multiple conversations. EX1007, ¶¶28, 31; EX1002, ¶124. For example, the system may store name data about “Matsui” in a prior conversation and use it to identify a conversation subject of a later conversation. *Id.* A POSITA would thus 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, and is “*long-term knowledge*” because it is user-centric and persists across multiple conversations. EX1007, ¶¶28, 31; EX1002, ¶125.

The stored name data includes information accumulated “prior to the “*predetermined time period*.” See EX1007, ¶¶28, 31, 77-84; EX1002, ¶¶124, 126. At least some of the stored name data is from conversations occurring “*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., EX1007, ¶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”); EX1002, ¶126. Therefore, Matsuda’s long-term knowledge is accumulated *“based on one or more natural language utterances received prior to the predetermined time period.”* EX1002, ¶126.

A POSITA would have been motivated to combine Kennewick and Matsuda, and the combination discloses *“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.”* EX1002, ¶127. Kennewick and Matsuda are analogous art; both describe dialog-based interaction systems that accumulate long-term knowledge and use the knowledge to provide more natural computer-generated responses to human natural language input. *See, e.g.,* EX1004, ¶11; EX1007, ¶84; EX1002, ¶127.

A POSITA would have found it obvious to incorporate Matsuda’s disclosure for accumulating long-term knowledge about conversation subjects into Kennewick’s system. EX1002, ¶128. 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, 116; EX1002, ¶128. Likewise, 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, 143; EX1002, ¶128. Accordingly, a POSITA would have been motivated to incorporate Matsuda's disclosure related to accumulating 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. EX1002, ¶128. A POSITA would have likewise found it obvious for other types of 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 be accumulated based on natural language utterances received prior to a "predetermined time period," as taught by Matsuda. *Id.*; *see also* EX1007, ¶¶108, 113 (embodiments set forth in Matsuda are not limited to person names).

The combination of Kennewick and Matsuda is a combination of prior art elements according to known methods to yield predictable results. EX1002, ¶129. The combination of Kennewick and Matsuda is also an application of a known technique, e.g., Matsuda's technique for accumulating and using long-term knowledge, to a known system, e.g., Kennewick's speech-based conversational interface that uses long-term knowledge including user profile information, that is ready for improvement. EX1002, ¶129. 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. EX1002, ¶129.

[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”

Kennewick in combination with Cooper discloses and renders obvious limitation [1.g]. EX1002, ¶¶130-138. As discussed above for limitations [1.a], [1.e], and [1.f], Kennewick and Matsuda disclose receiving a user input comprising a natural language utterance, and accumulating “*short-term knowledge*” and “*long-term knowledge*.” EX1002, ¶130. Kennewick further discloses using different personalities to generate responses exhibiting different personality traits such as “sarcasm, humor, irritation, and sympathy.” EX1004, ¶141; EX1002, ¶131. This 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.” EX1004, ¶24; EX1002, ¶131.

Kennewick does not explicitly disclose that its computer system “*identifies]...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, which describes a “virtual assistant [that] adapts to the user in many different ways based on” natural language utterances received from the user.

EX1005, 2:56-64, Abstract; *see also id.*, 6:35-37, 7:50-53, 46:58-59; EX1002, ¶¶132-133. The virtual assistant “*identifies]...a manner in which the natural language utterance was spoken,*” by determining “whether words associated with” certain types of discourse, e.g., “polite discourse,” are “included in input received from the user.” EX1005, 43:55-59, 2:26-30; EX1002, ¶134. Cooper also explains that the virtual assistant may “automatically determine the user’s emotional state, calm or angry, for example,” based on “information about the user’s emotion” contained in the user input, such as “information about the user’s voice volume, word choice and speech rate.” EX1005, 43:62-67; EX1002, ¶134. Cooper thus “*identifies]...a manner in which the natural language utterance was spoken*” by determining whether the utterance included words associated with polite discourse and/or by determining whether the user’s word choice conveys an emotional state such as being calm or angry. EX1005, 43:55-67; EX1002, ¶134. Cooper further discloses saving information about the user’s emotional state for use in future sessions. EX1005, 43:62-44-6; EX1002, ¶136.

It would have been 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.*” EX1002, ¶135. As discussed above, Kennewick’s computer system accumulates the short-term knowledge from the dialog with the user. EX1004, ¶¶23, 25, 29, 31, 37, 155, 157,

160; EX1002, ¶135. 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; EX1002, ¶135. 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 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." *See, e.g.*, EX1004, ¶155; EX1002, ¶135.

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.***" EX1002, ¶136. 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. EX1004, ¶¶31, 143; EX1002, ¶136. The user profile includes, 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, ¶143; EX1002, ¶136. Cooper also teaches saving information from past sessions, specifically user emotional state information, "for use in future sessions." EX1005, 43:55-44:6;

EX1002, ¶136. 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 in Kennewick's computer system modified by Cooper to further identify the manner in which the natural language utterance was spoken based on the long-term knowledge, 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; EX1002, ¶136.

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.*” EX1002, ¶137. As discussed above, Cooper identifies a user's manner of speaking based on the word choice of the user's utterances in the current conversation (short-term knowledge) and saved “information obtained about the user's emotional state” from past sessions (long-term knowledge) to identify a 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; EX1002, ¶137. Likewise, Kennewick's system “makes significant use of” short- and long-term knowledge “to achieve a natural environment for one or more

users making queries or commands,” including selecting personality characteristics for responding to the user. EX1004, ¶¶11, 35, 178; EX1002, ¶137. A POSITA 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 from past sessions (long-term knowledge) to identify the manner in which the utterance was spoken. EX1002, ¶137. 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. EX1002, ¶137.

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. EX1002, ¶138. The combination is also a predictable use of known techniques for improving voice-based interactions with computers. EX1002, ¶138. 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. EX1002, ¶138.

[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.”

Kennewick in combination with Cooper discloses limitation [1.h]. EX1002, ¶¶139-147. As discussed for the preamble and limitations [1.d] and [1.g], Kennewick’s computer system determines the “*interpretation*” of the natural language utterance, and in combination with Cooper discloses identifying the “*manner in which the natural language utterance was spoken.*” EX1004, ¶¶123, 155, 160, 162-63, 165; EX1005, 43:51-44:6; 2:26-30; EX1002, ¶139. Kennewick in combination with Cooper also discloses that generated responses are “*adapted based on a user’s manner of speaking*” as discussed for the preamble. EX1002, ¶¶83-90, 139.

Kennewick further discloses its computer system “*generating [] a response to the natural language utterance based on the interpretation.*” See, e.g., EX1004, ¶14; EX1002, ¶140. Kennewick’s computer system invokes an agent that works with the parser to interpret the utterance based on the context. The agent generates an appropriate response to user’s natural language utterance for output in a suitable presentation format. EX1004, ¶¶17, 124, 132, 162, 178; EX1002, ¶141. The agent may “format the response into the markup format used by the text to speech engine 124,” which “creates the required utterances” that are “output by the speaker 136.” EX1004, ¶¶23-25, 32-33, 124, 178-82; EX1002, ¶141. Thus, Kennewick discloses

and renders obvious “*generating, by the computer system, a response to the natural language utterance based on the interpretation.*” EX1002, ¶142.

Kennewick further discloses generating the response based on a “simulated personality.” EX1004, ¶¶24, 178-180; EX1002, ¶143. However, Kennewick does not explicitly disclose generating the response based on “*the identified manner in which the natural language utterance was spoken.*” This aspect is taught by Cooper. EX1002, ¶¶143-144.

As discussed for the preamble, Cooper describes a “virtual assistant [that] adapts to the user in many different ways based on the [received] input.” EX1005, 2:26-30, 2:56-64, Abstract; EX1002, ¶¶ 83-84. Cooper’s virtual assistant generates a response based on “*the identified manner in which the natural language utterance was spoken.*” For example, if the user uses “polite discourse,” the virtual assistant generates a polite response. EX1005, 43:48-61; EX1002, ¶144. If the user is angry, the virtual assistant generates a submissive response. EX1005, 43:62-44:4; EX1002, ¶144.

A POSITA would have found it obvious to incorporate Cooper’s teaching to generate a response based on the user’s manner into Kennewick’s agent-based system because it would improve the naturalness and user-friendliness of voice-based interactions. EX1002, ¶145. Kennewick describes “creat[ing]” and “format[ing] [a] response” to include “simulated personality characteristics”

selected based on various factors, while Cooper discloses generating a response that complements the user's input manner. EX1004, ¶¶24, 35, 124, 178; EX1005, 43:48–44:4; EX1002, ¶145. Combining the references so that Kennewick selects a response “personality” based on the identified manner as taught by Cooper, would have been a straightforward application of known techniques to improve the user experience. EX1004, ¶¶24, 178; EX1002, ¶145.

Such a combination would yield predictable results, e.g., a polite response to a polite user input. EX1002, ¶146. EX1004, ¶24. A POSITA would have reasonably expected the combined system to be successful because these technologies were well-understood. EX1002, ¶146. 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.*” EX1002, ¶147.

B. 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.”

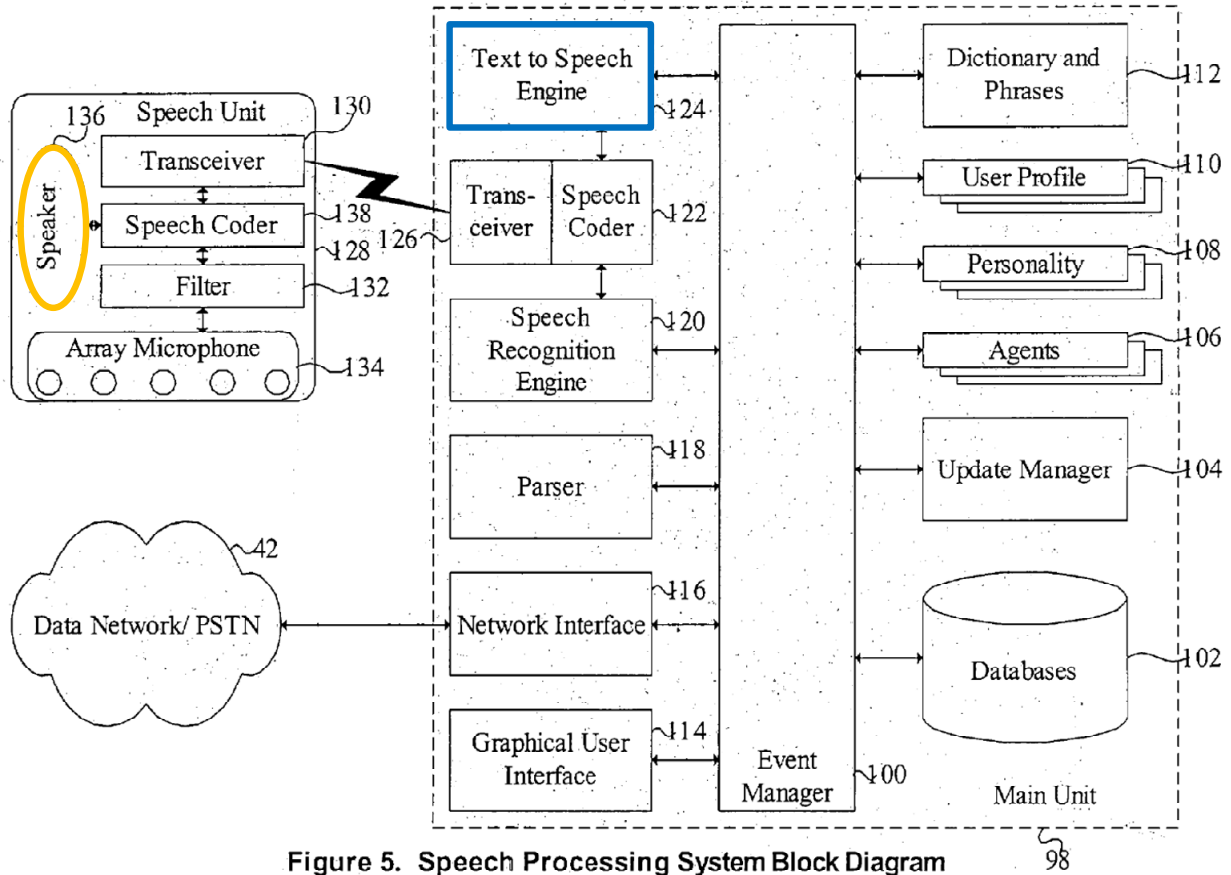
As discussed above for [1.pre] and [1.g], Kennewick in combination with Cooper discloses identifying the “*manner in which the natural language utterance was spoken,*” such as by determining whether the user's input includes “words associated with polite discourse” or contains information indicative of the user being

“calm or angry.” EX1005, 43:55-67; EX1002, ¶¶148-151. Cooper further discloses that the identified manner “*includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.*” EX1002, ¶149. For example, Cooper discloses that the virtual assistant determines whether the “input received from the user” contains “words associated with polite discourse” or contains “information about the user’s emotion” such as “the user’s voice volume, word choice and speech rate.” EX1005, 43:55-67; EX1002, ¶¶149-150. This amounts to the system identifying “*an indication of at least one of...word use, and/or jargon*” in the received input corresponding to the manner. EX1002, ¶150. In addition, Cooper’s virtual assistant detecting “information about the user’s...speech rate” in the received input is the system identifying “*an indication of at least one of...pace, [and/or] timing*” in the received input corresponding to the manner. EX1002, ¶150.

- C. **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.”**

Kennewick and Cooper both disclose wherein “*the response comprises a voice response.*” EX1002, ¶¶152-159. Kennewick’s computer system invokes an agent which generates and outputs an appropriate response to the user. EX1004, ¶¶124, 162; EX1002, ¶153. It sends the response to a [text to speech engine](#) which

creates “the required utterances” that are output through a **speaker**. The response is thus “*a voice response*.” EX1004, ¶124; EX1002, ¶153.



EX1004, Figure 5. Similarly, Cooper’s virtual assistant generates a “*voice response*” using “text-to-speech technology,” which is communicated to the user “in spoken English.” EX1005, 6:35-37, 8:5-7; EX1002, ¶¶155-156.

Further, as discussed for [1.h], Kennewick in combination with Cooper discloses generating the “*response based on the identified manner*.” EX1002, ¶154. Because the response may be a voice response, a POSITA would have

understood that Kennewick's computer system, as modified by Cooper, discloses ***“generating the voice response based on the identified manner.”*** EX1002, ¶154.

Cooper further discloses that when generating the voice response, the virtual assistant ***“var[ies] one or more of...word use, and/or jargon of the voice response,”*** by using words associated with a selected type of discourse, e.g., “polite discourse” or “submissive discourse,” ***“based on the identified manner,”*** e.g., whether the user is polite or angry. EX1005, 43:48-61, 43:62-44:4; EX1002, ¶¶157-158. 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 to advance 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; EX1002, ¶159. 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,” and present responses that appear less “rigidly formatted or mechanical.” EX1004, ¶24; EX1005, Abstract; EX1002, ¶159.

D. Claim 4

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

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

A POSITA would have found it obvious that Kennewick’s “***computer system***” “***obtain[s]...grammatical rules***” to format responses to user utterances. EX1002, ¶¶160-163. Kennewick’s computer system invokes an agent which generates an appropriate response to the user’s natural language utterance and “format[s] that response for presentation.” EX1004, ¶¶17, 124, 132, 162, 178; EX1002, ¶161.

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. EX1002, ¶¶162-163. For example, Kennewick discloses using “presentation templates” to generate “natural language speech response[s]” so as to “enhance[e] the understandability [of the responses] to the user.” EX1004, ¶¶14, 23, 179, 181; EX1002, ¶162. A POSITA would have understood that a “template” specifying the format of a “natural language speech response” would specify rules or guidelines for how the response is to be generated. EX1002, ¶162. For a speech response, this would include rules relating to the grammar and syntax of the response, so that the generated response will be in a grammatically correct natural language format that the user can understand. *See id.* For example, Huang, a well-known textbook, described that

conventional systems obtained grammatical rules in this manner. EX1011, 913 (describing use of “dialog grammars” with “syntactic grammar rules [that] state constraints on grammatically acceptable strings”), 924 (illustrating use of “Grammar” and “Rules for Speech Synthesis” to generate synthesized speech responses); EX1002, ¶162. A POSITA would have found it obvious for Kennewick’s system to obtain grammatical rules when preparing a response, so that the generated “natural language speech response” is understandable to the user. EX1002, ¶¶162-163.

[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.”

As discussed for limitation [1.h], Kennewick in combination with Cooper discloses “*generating the response based on the identified manner in which the natural language utterance was spoken.*” EX1002, ¶¶164-167. Kennewick also discloses “*using the obtained... grammatical rules to generate sentences for use as response sets to cooperate with the user.*” EX1002, ¶165.

As discussed above, a POSITA would have understood that Kennewick’s computer system “*obtain[s]...grammatical rules*” to generate and format a response. EX1002, ¶165. In other words, Kennewick’s system “*us[es] the obtained...grammatical rules*” to generate and format the response. EX1004, ¶14; EX1002, ¶165. In addition, Kennewick discloses that the generated response may

be a “natural language speech response,” which a POSITA would have found obvious to include “*sentences*,” so that the response can be conveyed to the user in a “natural language.” See EX1004, ¶¶14, 17, 36; EX1002, ¶165. It would have been obvious that the system generates multiple sentences for the response in cases where the information to be conveyed in the response requires more than a single sentence. See EX1004, ¶¶17, 36; EX1002, ¶165.

Kennewick further discloses using “*response sets*” to present information to the user. See EX1004, ¶¶25, 36; EX1002, ¶166. For example, the system may divide a “long set of information” into subsets, to be presented to the user through a set of one or more responses. EX1004, ¶¶14, 25, 36; EX1002, ¶166. A POSITA would have found it obvious for the sentences generated by the system to be used as response sets, where the system communicates the sentences to the user through a set of responses. EX1002, ¶166.

Kennewick further discloses that the sentences generated by the system “*cooperate with the user*.” EX1002, ¶167. For example, Kennewick discloses generating and providing “a satisfactory response to [the user’s] question [or] command.” See, e.g., EX1004, ¶¶31-32, 170-178; EX1002, ¶167. The sentences used as response sets thus cooperate with the user by providing information to answer the user’s questions. See *id.* In addition, Kennewick further discloses formatting the response to be “more natural” to the user, which “*cooperate[s] with*

the user” because a response transformed to be more understandable and natural to the user is a response that cooperates with the user. EX1004, ¶178; EX1002, ¶167. 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.*” EX1002, ¶167.

E. Claim 5

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

Kennewick discloses accumulating “*long-term knowledge*” in a user profile as discussed for limitation [1.f]. EX1004, ¶¶11, 23, 31, 143; EX1002, ¶¶118-120. The long-term knowledge in a user profile “*is associated with a [] user*” because it is specific to an individual user. EX1004, ¶¶11, 23, 31, 143; EX1002, ¶¶168-169. A POSITA would have found it obvious that the “user” may be “*a first user*” among multiple users because the computer system of Kennewick is designed to work with “one or more users.” EX1002, ¶169.

[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.”

Kennewick discloses limitation [5.a]. EX1002, ¶¶170-171. As discussed for limitation [5.pre], Kennewick’s “*computer system*” creates and stores

(“*generat[es]*”) for the user (e.g., “*first user*”), the user profile (“*a profile*”) by accumulating “*long-term knowledge*” about the user. EX1004, ¶¶31, 143; EX1002, ¶170. The system generates the profile and keeps it “updated automatically” by adding additional long-term knowledge, e.g., histories, over time “as the user uses the system.” EX1004, ¶31; EX1002, ¶170. This meets the claimed “*generating, by the computer system, a profile associated with the first user based on the long-term knowledge,*” and is consistent with the ’699 patent’s description that “long-term knowledge may accumulate over time to build user profiles.” EX1001, 14:6-7; EX1002, ¶170. Further, Kennewick discloses that “*the context for the natural language utterance is determined based further on the profile associated with the first user,*” because Kennewick explicitly discloses that the parser “determine[s] a context for an utterance” based on the “user profile 110.” EX1004, ¶160; EX1002, ¶171.

F. Claim 6: “The method of claim 1, wherein generating the response based on the identified manner comprises: adapting, by the computer system, the response based on a response format associated with the identified manner.”

Kennewick and Cooper disclose “*generating the response based on the identified manner*” as discussed for limitation [1.h]. EX1002, ¶172. Kennewick further discloses that the “*generating*” comprises “*adapting...the response.*” Kennewick’s system adapts the response using a selected template and other information. EX1004, ¶¶178-180; EX1002, ¶¶172-173. For example, “the domain

agent 156 may select the presentation template, determine order of presentation for tokens and determine variable substitutions and transformations [for the response] using probabilistic or fuzzy set decision methods” where the “variable substitutions and transformations may be derived from a number of sources.” EX1004, ¶¶179-180.

Kennewick does not explicitly disclose that the adapting is “*based on a response format associated with the identified manner.*” EX1002, ¶¶174-175. This aspect is taught by Cooper, which discloses adapting the response based on the user’s identified manner as discussed for limitation [1.g]. EX1005, 43:55-67, 2:26-30; EX1002, ¶¶133-134. For example, Cooper discloses adapting the response by including words associated with polite or submissive discourse. EX1005, 43:51-55, 44:1-4; EX1002, ¶175. Cooper’s adaptations are “*based on a response format associated with the identified manner*” because they use a particular response format (e.g., words associated with polite or submissive discourse) selected based on the user’s identified manner. EX1002, ¶¶174-175.

A POSITA would have found it obvious to incorporate Cooper’s teachings into Kennewick’s system. EX1002, ¶176. Kennewick discloses adapting the response based on various factors. EX1004, ¶¶179-180; EX1002, ¶176. Cooper discloses that the response may be adapted using a response format associated with the user’s identified manner. EX1005, 43:55-67, 2:26-30; EX1002, ¶176. 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]. EX1002, ¶176.

G. Claim 7

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

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

Kennewick discloses these limitations. EX1002, ¶¶177-179. Kennewick explicitly discloses that the “*computer system*” “*provid[es]...the natural language utterance as an input to a speech recognition engine*” in order to “*recogniz[e] the one or more words or phrases from the natural language utterance*” for the reasons discussed for limitation [1.b]. *See, e.g.*, EX1004, ¶155 (“The output of the speech coder 122 may be fed to the speech recognition engine 120.”); EX1002, ¶179.

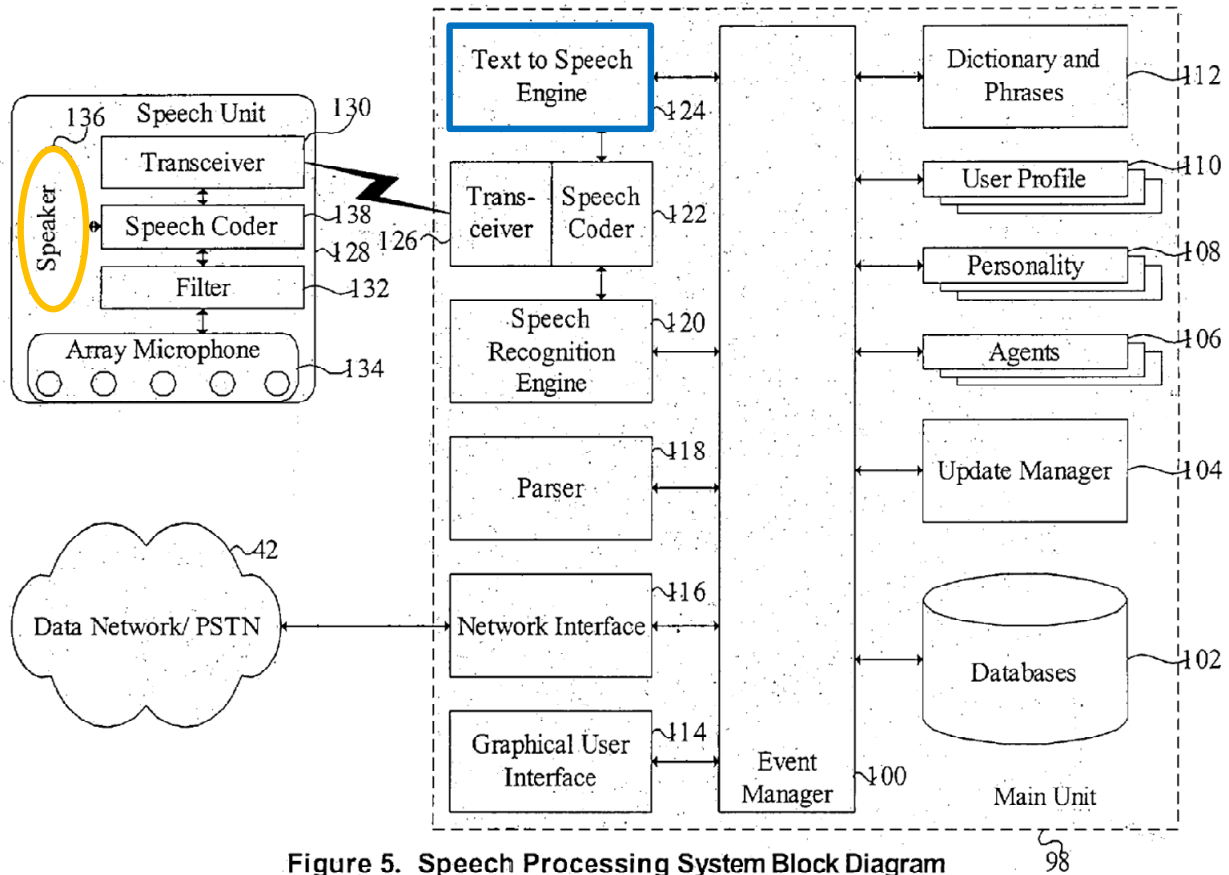
[7.b] “obtaining, by the computer system, the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.”

Kennewick further discloses that the speech recognition engine 120 outputs recognized “words and phrases,” which, as discussed in relation to limitation [1.c], are passed “to the parser 118 for interpretation.” EX1004, ¶¶123, 155; EX1002, ¶180. Kennewick’s parser 118 therefore “*obtain[es]...the one or more words or*

phrases recognized from the natural language utterance as an output of the speech recognition engine.” EX1002, ¶181.

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

As discussed for limitation [1.h], Kennewick in combination with Cooper discloses generating the “*response to the natural language utterance.*” EX1002, ¶¶182-183. Kennewick further discloses that the “*computer system*” causes the generated response “*to be provided to the user.*” It “create[s] a response string” that is sent “to the *text to speech engine 124*” and “output by the *speaker 136*” so the user can hear it. EX1004, ¶124, Figure 3; EX1002, ¶184.



EX1004, Figure 5; EX1002, ¶184.

- I. **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.”**

Kennewick in combination with Cooper discloses claim 9 for the same reasons provided for claim 3. EX1002, ¶185.

J. Claim 10: “The method of claim 1, further comprising: adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.”

As discussed for limitation [1.h], Kennewick in combination with Cooper discloses generating “*the response*” to the natural language utterance. EX1002, ¶¶186-187. Kennewick further discloses “*adapting the response to have a personality by varying word use*,” by selecting a personality and formatting (adapting) responses to have that personality. EX1004, ¶¶24, 35, 141, 180; EX1002, ¶188. For example, responses “that may be upsetting to the user” may be “presented in a sympathetic manner,” while “information requiring immediate action or announcing a safety problem can be delivered with a definite and authoritative personality.” EX1004, ¶24; EX1002, ¶188. These adaptations involve “*varying word use*.” EX1004, ¶¶178 (“The domain agent may vary the order of presentation of tokens and the exact terminology used to create a more natural response to the user.”), 180 (using “preferred terminology” based on “presentation personality”); EX1002, ¶188.

Cooper also discloses “*adapting the response to have a personality by varying word use*.” EX1002, ¶189. As discussed for limitation [6.a], Cooper adapts the response to have a polite or submissive personality by variably including words associated with polite or submissive discourse. EX1005, 43:51-55, 44:1-4; EX1002, ¶189. These adaptations to the response “*model a conversation*” because the words

in the response are varied to imitate the user's manner of speaking. EX1005, 43:48-44:6; EX1002, ¶190. Cooper's disclosure is similar to that of the '699 patent, which describes modeling a conversation by imitating a user's input word choices. EX1001, 18:57-58, 19:1-19; EX1002, ¶190.

A POSITA would have found it obvious to incorporate Cooper's teachings of "*modell[ing] a conversation*" by "*adapting the response to have a personality by varying word use*" into Kennewick's system. EX1002, ¶191. For example, Kennewick discloses adapting responses to use "preferred terminology" associated with different "personality characteristics." EX1004, ¶¶24, 180; EX1002, ¶191. Cooper further discloses modeling a conversation by varying the word use of responses to exhibit specific personality traits (e.g., politeness, submissiveness) to match or complement the manner of the user's inputs. *See* EX1005, 43:48-44:6; EX1002, ¶191. By modifying Kennewick's system to adapt the system's personality based on how the user speaks to the system as taught by Cooper, the system can better "simulate the behavioral characteristics of real humans" in a conversation, and provide "responses to user's questions and commands seem more natural." EX1004, ¶¶24, 35; *see also* EX1005, Abstract (system "adapts to the user...based on the input [it] receives"); EX1002, ¶191. 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.” EX1002, ¶191.

K. 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.”

As discussed 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 determined based on the identified context. EX1002, ¶¶192-193. This response is also “*sensitive to context*” because it may vary based on the context. EX1002, ¶193.

Further, as discussed 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*.” EX1002, ¶¶194-195. The response is also generated sensitive to “*words uttered by the user in one or more prior natural language utterances*.” EX1002, ¶196. 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 involving prior natural language utterances. EX1004, ¶¶31, 143; EX1002, ¶196. Kennewick’s disclosure of generating a response based this historical information demonstrates that the response is

“sensitive to...words uttered by the user in one or more prior natural language utterances.” EX1002, ¶196.

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.”* EX1002, ¶197. For example, when interpreting the user’s utterance, Kennewick’s system determines “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; EX1002, ¶197. In addition, as discussed above, the generated response is sensitive to the contents of the user profile. The user profile contains information “manually added or updated by the user.” EX1004, ¶31; EX1002, ¶197. A POSITA would have understood that information “explicitly supplied by the user” in an utterance or “manually added or updated by the user” into the profile is information of *“what the user already knows about a topic.”* EX1002, ¶197. A POSITA would have understood that Kennewick’s computer system generating a response based on this information demonstrates that the response is *“sensitive to...what the user already knows about a topic.”* EX1004, ¶31; EX1002, ¶197.

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; EX1002, ¶198. Kennewick teaches that its system makes “maximum use of context, prior information,...and user specific profile data” (which as explained above encompasses the various pieces of claimed information) to interpret the user’s utterance and generate the response. EX1004, ¶118; EX1002, ¶198. A POSITA would therefore have understood that the system will consider all of this information when generating the response. EX1002, ¶198. The generated response will thus be “*sensitive to*” each of these pieces of information. *Id.* 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.*” EX1002, ¶198.

L. 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:”

The combination of Kennewick, Cooper, and Matsuda renders claim 12 unpatentable for the same reasons as claim 1. EX1002, ¶¶199-200. Preamble

[12.pre] and limitation [12.a] are substantively identical to preamble [1.pre]. EX1002, ¶200. 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. EX1002, ¶200. 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. EX1002, ¶200. The “*computer system*” implementing the method of claim 1 is equivalent to the “*system*” of claim 12 and obvious for the same reasons. EX1002, ¶200. The remaining limitations of claim 12 are substantively identical to the corresponding limitations of claim 1 and obvious for the same reasons. EX1002, ¶200.

M. Claims 13-22

Each of dependent claims 13-22 is substantively identical to corresponding claims 2-11, respectively, containing only minor differences to reflect that independent claim 12 recites a system instead of a method. EX1002, ¶¶201-202. Accordingly, claims 13-22 are unpatentable for the same reasons as provided for claim 12 and the respective corresponding claims. EX1002, ¶202.

VI. CONCLUSION

For these reasons, Petitioner requests *inter partes* review and cancellation of the challenged claims.

VII. STANDING

Petitioner certifies that Petitioner is not barred or estopped from requesting this *inter partes* review and the '699 patent is IPR eligible.

VIII. MANDATORY NOTICES AND FEES

A. Real Party-in-Interest

Petitioner is the real party-in-interest. No other party directed, controlled, or funded this IPR proceeding.

B. Related Matters

Patent Owner asserts the '699 patent against Petitioner in *VB Assets, LLC v. Amazon.com Services, LLC*, No. 1:24-cv-00839 (District of Delaware, July 18, 2024) (the "parallel litigation").

Petitioner has filed a petition for *inter partes* review of U.S. Patent No. 10,297,249, which is related to the '699 patent.

Petitioner is aware of the following additional related matters involving the '699 patent and/or related patents:

Case Caption	Forum	Patents
<i>VB Assets LLC v. SoundHound AI Inc.</i> , Case No. 1:24-cv-01279	DDE	7,818,176 10,755,699 11,087,385

Petition for *Inter Partes* Review
U.S. Patent No. 10,755,699

Case Caption	Forum	Patents
		11,222,626 10,297,249 9,269,097 9,502,025 8,886,536 8,073,681
<i>VB Assets LLC v. Amazon.com Services LLC et al.</i> , Case No. 1:19-cv-01410	DDE	7,818,176 9,626,703 9,015,049 9,269,097 8,886,536 8,073,681
<i>VB Assets LLC v. Apple Inc.</i> , Case No. 1:24-cv-01368	DDE	7,818,176 8,073,681 10,297,249 10,510,341 8,515,765 8,886,536 9,269,097
<i>VB Assets LLC v. Samsung Electronics Co., Ltd. et al.</i> , Case No. 2:24-cv-00828	E.D. Tex.	8,073,681 10,755,699 10,510,341 8,515,765 8,886,536 7,818,176
<i>VB Assets LLC v. Amazon.com Services LLC</i> , Case No. 25-1142	Fed. Cir.	7,818,176 9,626,703 9,015,049 9,269,097 8,886,536 8,073,681
<i>VB Assets LLC v. Amazon.com Services LLC</i> , Case No. 25-1357	Fed. Cir.	7,818,176 8,073,681 9,269,097

Case Caption	Forum	Patents
<i>Amazon.com Services LLC v. VB Assets LLC</i> , Case No. 25-1113	Fed. Cir.	7,818,176 9,626,703 8,073,681 9,269,097
IPR2020-01346	PTAB	9,015,049
IPR2025-00868	PTAB	8,515,765
IPR2025-00866	PTAB	8,073,681
IPR2020-01367	PTAB	8,073,681
IPR2025-00870	PTAB	10,755,699
IPR2025-00871	PTAB	10,510,341

C. Lead and Backup Counsel

Pursuant to 37 C.F.R. §42.8(b)(3) and 42.10(a), Petitioner designates J. David Hadden, Reg. No. 40,629 as lead counsel and Brian M. Hoffman, Reg. No. 39,713, Saina Shamilov, Reg. No. 48,266, Dargaye Churnet, Reg. No. 71,288, Eric Zhou, Reg. No. 68,842, and Jwalant R. Dholakia, Reg. No. 77,732, as back-up counsel, each of Fenwick & West LLP.

D. Service Information

Petitioner consents to service by electronic mail at:

VBAssets-IPR@fenwick.com.

Petitioner's counsel may also be served by mail or hand delivery at Fenwick & West LLP, 801 California St, Mountain View, CA 94041. Petitioner's counsel may be reached by telephone at (650) 988-8500.

E. Fees

The Office is authorized to charge fees for this Petition to Deposit Account 19-2555.

Dated: August 4, 2025

FENWICK & WEST LLP

/J. David Hadden/

J. David Hadden
Reg. No. 40,629

Attorneys for Petitioner
Amazon.com Services LLC

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. §42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. § 42.8, certificates of service or word count, or appendix of exhibits, contains 11,408 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: August 4, 2025

FENWICK & WEST LLP

/J. David Hadden/

J. David Hadden

Reg. No. 40,629

Attorneys for Petitioner
Amazon.com Services LLC

**CERTIFICATE OF SERVICE ON PATENT OWNER
UNDER 37 C.F.R. § 42.105**

I hereby certify, pursuant to 37 C.F.R. §§ 42.6(e) and 42.105, that a complete copy of this Petition for *Inter Partes* Review including all exhibits are being served via Federal Express on August 4, 2025, upon Patent Owner by serving the correspondence address of record with the USPTO as follows:

James Gatto
Daniel Yannuzzi
David Heisey
Sheppard Mullin Richter & Hampton
650 Town Center Drive, 10th Floor
Costa Mesa, CA 92626
Patent Owner's correspondence

address of record for US Patent 10,755,699

Dated: August 4, 2025

FENWICK & WEST LLP

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioner
Amazon.com Services, Inc.