

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: IPR2025-01365

U.S. Patent No. 10,297,249

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 10,297,249**

TABLE OF CONTENTS

	Page
I. OVERVIEW OF THE '249 PATENT	1
A. Alleged Invention	1
B. Priority Date	2
II. IDENTIFICATION OF CHALLENGE	2
A. Statutory Grounds.....	2
B. Prior Art.....	3
1. Kennewick (EX1004)	3
2. Oviatt (EX1005).....	4
3. Fujimoto (EX1007, EX1008).....	5
4. Shimomura (EX1014).....	6
5. Matsuda (EX1015, EX1016).....	7
6. Cooper (EX1017).....	7
III. LEVEL OF ORDINARY SKILL IN THE ART	8
IV. CLAIM CONSTRUCTION	8
V. GROUND 1: KENNEWICK, OVIATT, AND FUJIMOTO	9
A. Claim 1	9
[1.pre] “A computer-implemented method of facilitating natural language system responses using short-term knowledge generated based on one or more prior multi-modal device interactions, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:”	9

[1.a]	“receiving, by the computer system during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;”	16
[1.b]	“receiving, by the computer system, a second voice input comprising the first natural language utterance via a second input device;”	17
[1.c]	“comparing, by the computer system, the first voice input with the second voice input;”	18
[1.d]	“filtering, by the computer system, sound from the first voice input and the second voice input based on the comparison;”	22
[1.e]	“obtaining, by the computer system during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input;”	23
[1.f]	“generating, by the computer system, the short-term knowledge based on at least the first voice input and the first non-voice input;”	28
[1.g]	“determining, by the computer system, based on the short-term knowledge, a first context for the first natural language utterance;”	30
[1.h]	“determining, by the computer system, based on the first context, an interpretation of the first natural language utterance; and”	31
[1.i]	“generating, by the computer system, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.” ...	32
B.	Claim 6	33
	“The method of claim 1, wherein determining the interpretation of the first natural language utterance	

	comprises determining, based on the first context, an interpretation of one or more recognized words of the first natural language utterance.”	33
C.	Claim 9	34
	[9.pre] “The method of claim 1, the method further comprising:”	34
	[9.a] “accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received during conversations occurring prior to the first conversation, wherein the first context is determined based on the short-term knowledge and the long-term knowledge.”	34
D.	Claim 12	36
	“The method of claim 1, wherein the second input device is included with the first input device in a single device.”	36
E.	Claim 13	36
	“The method of claim 1, wherein generating the first response comprises generating, by the computer system, a non-voice response in a graphical user interface and/or a voice response via a speaker.”	36
F.	Claim 14	38
	[14.pre] “The method of claim 1, wherein the first natural language utterance relates to a request and the first response is intended to be responsive to the request, the method further comprising:”	38
	[14.a] “receiving, by the computer system, a second natural language utterance in response to the first response, wherein the second natural language utterance comprises a clarification related to the interpretation of the first natural language utterance; and”	39

	[14.b] “responsive to receipt of the second natural language utterance, generating, by the computer system, a response to the second natural language utterance based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance.”...39	
G.	Claim 15	40
	[15.pre] “The method of claim 1, wherein the first non-voice input is received via a non-voice input device comprising a touch screen device, the method further comprising:”	40
	[15.a] “receiving, by the computer system, information indicating a location on the touch screen device that received the first non-voice input.”	42
H.	Claim 16	43
	[16.pre] “A system for facilitating natural language system responses via short-term knowledge generated based on one or more prior multi-modal device interactions, the system comprising:”	43
	[16.a] “one or more physical processors programmed with one or more computer program instructions which, when executed, cause the one or more physical processors to:”	43
I.	Claims 21 and 25-28.....	44
VI.	GROUND 2: KENNEWICK, OVIATT, FUJIMOTO, AND SHIMOMURA	44
A.	Claim 2	44
	[2.pre] “The method of claim 1, the method further comprising:”	44
	[2.a] “receiving, at the computer system, a second natural language utterance via the first input device;”	44

[2.b]	“determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to the first context;”	47
[2.c]	“determining, by the computer system, responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge;”	51
[2.d]	“determining, by the computer system, based on the second context, an interpretation of the second natural language utterance; and”	52
[2.e]	“generating, by the computer system, based on the interpretation of the second natural language utterance, a response to the second natural language utterance.” ...	52
B.	Claim 3	53
[3.pre]	“The method of claim 1, the method further comprising:”	53
[3.a]	“tracking, by the computer system, contexts identified for multiple consecutive natural language user utterances to generate a context stack, the context stack including the contexts identified for the multiple consecutive natural language user utterances;”	53
[3.b]	“receiving, at the computer system, a second natural language utterance via the first input device;”	56
[3.c]	“determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.”	56
C.	Claim 4	58
[4.pre]	“The method of claim 3, wherein determining whether the second natural language utterance corresponds to	

	one or more individual ones of the contexts included in the context stack includes:”	58
	[4.a] “determining whether the second natural language utterance corresponds to a most recent context included in the context stack; and”	58
	[4.b] “responsive to the second natural language utterance not corresponding to the most recent context, determining whether the second natural language utterance corresponds to a second most recent context included in the context stack.”	59
D.	Claim 5	59
	[5.pre] “The method of claim 3, further comprising:”	59
	[5.a] “determining, by the computer system, responsive to the second natural language utterance not corresponding to one or more individual ones of the contexts included in the context stack, a second context for the second natural language utterance based on the short-term knowledge; and”	60
	[5.b] “determining, by the computer system, based on the second context, an interpretation of the second natural language utterance.”	61
E.	Claims 17-20	61
VII.	GROUND 3: KENNEWICK, OVIATT, FUJIMOTO, AND MATSUDA.....	62
A.	Claim 7	62
	[7.pre] “The method of claim 1, the method further comprising:”	62
	[7.a] “expiring, by the computer system, one or more items of short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation.”	62

B.	Claim 8	66
	[8.pre] “The method of claim 7, the method further comprising:”	66
	[8.a] “accumulating, by the computer system, long-term knowledge based on the one or more items of short-term knowledge that are expired.”	66
C.	Claim 22	67
	[22.pre] “The system of claim 16, wherein the one or more physical processors are caused to:”	67
	[22.a] “expire one or more items of accumulated short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation; and”	67
	[22.b] “accumulate long-term knowledge based on the one or more items of accumulated short-term knowledge that are expired, wherein the first context is determined based on short-term knowledge and the long-term knowledge.”	68
VIII.	GROUND 4: KENNEWICK, OVIATT, FUJIMOTO, AND COOPER.....	68
A.	Claim 10	68
	[10.pre] “The method of claim 1, further comprising:”	68
	[10.a] “determining, by the computer system, based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the first response is generated based on the determined manner and the interpretation of the first natural language utterance.”	68

B.	Claim 11	72
	“The method of claim 10, wherein the manner in which the first natural language utterance is spoken includes at least one of a tone, a pace, an inflection, or a timing.” ..	72
C.	Claims 23 and 24.....	73
IX.	CONCLUSION.....	73
X.	STANDING	73
XI.	MANDATORY NOTICES AND FEES	74
A.	Real Party-in-Interest	74
B.	Related Matters.....	74
C.	Lead and Backup Counsel.....	76
D.	Service Information.....	76
E.	Fees.....	76

TABLE OF AUTHORITIES

	Page(s)
Statutes	
35 U.S.C. §102	3

EXHIBIT LIST (37 C.F.R. § 42.63(e))

Exhibit	Description
1001	U.S. Patent No. 10,297,249 (“the ’249 patent”)
1002	Declaration of Padhraic Smyth, Ph.D.
1003	Prosecution File History for the ’249 patent
1004	U.S. Patent Application Publication No. 2004/0193420 A1 (“Kennewick”)
1005	“Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions” by Sharon Oviatt et al., <i>Human-Computer Interaction</i> (2000, Vol. 15) (“Oviatt”)
1006	Declaration of Rachel J. Watters on Authentication of Publication
1007	Japanese Patent Application No. 2004145025A
1008	Certified English Translation of Japanese Patent Application No. 2004145025A (“Fujimoto”)
1009	Asthana, A. & Krzyzanowski, P., “A Small Domain Communications System for Personalized Shopping Assistance,” Proceedings of 1994 International Conference on Personal Wireless Communications (pp.199-203), IEEE Press (1994, August), (“Asthana”)
1010	Lucente, M., “Conversational interfaces for e-commerce applications,” Communications of the ACM (Sept. 2000), 43(9), 59-61, (“Lucente”)
1011	Huang 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

Exhibit	Description
1014	U.S. Patent Application Publication No. 2001/0021909 A1 (“Shimomura”)
1015	Japanese Patent Application No. JP2004334591A
1016	Certified English Translation of Japanese Patent Application No. JP2004334591A (“Matsuda”)
1017	U.S. Patent No. 6,757,362 (“Cooper”)
1018	Curriculum Vitae of Padhraic Smyth, Ph.D.
1019	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
1020	Excerpt from Microsoft Computer Dictionary, Fifth Edition (2002)

CHART OF CLAIMS
[1.pre] A computer-implemented method of facilitating natural language system responses using short-term knowledge generated based on one or more prior multi-modal device interactions, the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:
[1.a] receiving, by the computer system during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;
[1.b] receiving, by the computer system, a second voice input comprising the first natural language utterance via a second input device;
[1.c] comparing, by the computer system, the first voice input with the second voice input;
[1.d] filtering, by the computer system, sound from the first voice input and the second voice input based on the comparison;
[1.e] obtaining, by the computer system during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input;
[1.f] generating, by the computer system, the short-term knowledge based on at least the first voice input and the first non-voice input;
[1.g] determining, by the computer system, based on the short-term knowledge, a first context for the first natural language utterance;
[1.h] determining, by the computer system, based on the first context, an interpretation of the first natural language utterance; and
[1.i] generating, by the computer system, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.
[2.pre] The method of claim 1, the method further comprising:

CHART OF CLAIMS	
[2.a]	receiving, at the computer system, a second natural language utterance via the first input device;
[2.b]	determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to the first context;
[2.c]	determining, by the computer system, responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge;
[2.d]	determining, by the computer system, based on the second context, an interpretation of the second natural language utterance; and
[2.e]	generating, by the computer system, based on the interpretation of the second natural language utterance, a response to the second natural language utterance.
[3.pre]	The method of claim 1, the method further comprising:
[3.a]	tracking, by the computer system, contexts identified for multiple consecutive natural language user utterances to generate a context stack, the context stack including the contexts identified for the multiple consecutive natural language user utterances;
[3.b]	receiving, at the computer system, a second natural language utterance via the first input device; and
[3.c]	determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.
[4.pre]	The method of claim 3, wherein determining whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack includes:
[4.a]	determining whether the second natural language utterance corresponds to a most recent context included in the context stack; and

CHART OF CLAIMS
[4.b] responsive to the second natural language utterance not corresponding to the most recent context, determining whether the second natural language utterance corresponds to a second most recent context included in the context stack.
[5.pre] The method of claim 3, further comprising:
[5.a] determining, by the computer system, responsive to the second natural language utterance not corresponding to one or more individual ones of the contexts included in the context stack, a second context for the second natural language utterance based on the short-term knowledge; and
[5.b] determining, by the computer system, based on the second context, an interpretation of the second natural language utterance.
[6] The method of claim 1, wherein determining the interpretation of the first natural language utterance comprises determining, based on the first context, an interpretation of one or more recognized words of the first natural language utterance.
[7.pre] The method of claim 1, the method further comprising:
[7.a] expiring, by the computer system, one or more items of short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation.
[8.pre] The method of claim 7, further comprising:
[8.a] accumulating, by the computer system, long-term knowledge based on the one or more items of short-term knowledge that are expired.
[9.pre] The method of claim 1, the method further comprising:
[9.a] accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received during conversations occurring prior to the first conversation, wherein the first context is determined based on the short-term knowledge and the long-term knowledge.
[10.pre] The method of claim 1, further comprising:

CHART OF CLAIMS
[10.a] determining, by the computer system, based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the first response is generated based on the determined manner and the interpretation of the first natural language utterance.
[11] The method of claim 10, wherein the manner in which the first natural language utterance is spoken includes at least one of a tone, a pace, an inflection, or a timing.
[12] The method of claim 1, wherein the second input device is included with the first input device in a single device.
[13] The method of claim 1, wherein generating the first response comprises generating, by the computer system, a non-voice response in a graphical user interface and/or a voice response via a speaker.
[14.pre] The method of claim 1, wherein the first natural language utterance relates to a request and the first response is intended to be responsive to the request, the method further comprising:
[14.a] receiving, by the computer system, a second natural language utterance in response to the first response, wherein the second natural language utterance comprises a clarification related to the interpretation of the first natural language utterance; and
[14.b] responsive to receipt of the second natural language utterance, generating, by the computer system, a response to the second natural language utterance based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance.
[15.pre] The method of claim 1, wherein the first non-voice input is received via a non-voice input device comprising a touch screen device, the method further comprising:
[15.a] receiving, by the computer system, information indicating a location on the touch screen device that received the first non-voice input.

CHART OF CLAIMS
[16.pre] A system for facilitating natural language system responses via short-term knowledge generated based on one or more prior multi-modal device interactions, the system comprising:
[16.a] one or more physical processors programmed with one or more computer program instructions which, when executed, cause the one or more physical processors to:
[16.a.i] receive, during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;
[16.a.ii] receive a second voice input comprising the first natural language utterance via a second input device;
[16.a.iii] compare the first voice input with the second voice input;
[16.a.iv] filter sound from the first voice input and the second voice input based on the comparison;
[16.a.v] obtain, during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input;
[16.a.vi] generate the short-term knowledge based on at least the first voice input and the first non-voice input;
[16.a.vii] determine, based on the short-term knowledge, a first context for the first natural language utterance;
[16.a.viii] determine, based on the first context, an interpretation of the first natural language utterance; and generate, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.
[17.pre] The system of claim 16, wherein the one or more physical processors are further caused to:
[17.a] receive a second natural language utterance via the first input device;
[17.b] determine, based on the short-term knowledge, whether the second natural language utterance corresponds to the first context;

CHART OF CLAIMS
[17.c] determine, responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge;
[17.d] determine, based on the second context, an interpretation of the second natural language utterance; and
[17.e] generate, based on the interpretation of the second natural language utterance, a response to the second natural language utterance.
[18.pre] The system of claim 16, wherein the one or more physical processors are further caused to:
[18.a] track contexts identified for multiple consecutive natural language user utterances to generate a context stack, the context stack including the contexts identified for the multiple consecutive natural language user utterances;
[18.b] receive a second natural language utterance via the first input device; and
[18.c] determine, based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.
[19.pre] The system of claim 18, wherein determining whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack includes:
[19.a] determining whether the second natural language utterance corresponds to a most recent context included in the context stack; and
[19.b] responsive to the second natural language utterance not corresponding to the most recent context, determining whether the second natural language utterance corresponds to a second most recent context included in the context stack.
[20.pre] The system of claim 18, wherein the one or more physical processors are caused to:
[20.a] determine, responsive to the second natural language utterance not corresponding to one or more individual ones of the contexts included in the context

CHART OF CLAIMS
stack, a second context for the second natural language utterance based on the short-term knowledge; and
[20.b] determine, based on the second context, an interpretation of the second natural language utterance.
[21] The system of claim 16, wherein determining the interpretation of the first natural language utterance comprises determining, based on the first context, an interpretation of one or more recognized words of the first natural language utterance.
[22.pre] The system of claim 16, wherein the one or more physical processors are caused to:
[22.a] expire one or more items of accumulated short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation; and
[22.b] accumulate long-term knowledge based on the one or more items of accumulated short-term knowledge that are expired, wherein the first context is determined based on short-term knowledge and the long-term knowledge.
[23.pre] The system of claim 16, wherein the one or more physical processors are caused to:
[23.a] determine, based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the first response is generated based on the determined manner and the interpretation of the first natural language utterance.
[24] The system of claim 23, wherein the manner in which the first natural language utterance is spoken includes at least one of a tone, a pace, an inflection, or a timing.
[25] The system of claim 16, wherein the second input device is included with the first input device in a single device.
[26.pre] The system of claim 16, wherein to generate the first response, the one or more physical processors are caused to:

CHART OF CLAIMS	
[26.a]	generate a non-voice response in a graphical user interface and/or a voice response via a speaker.
[27.pre]	The system of claim 16, wherein the first natural language utterance relates to a request and the first response is intended to be responsive to the request, wherein the one or more physical processors are further caused to:
[27.a]	receive a second natural language utterance in response to the first response, wherein the second natural language utterance comprises a clarification related to the interpretation of the first natural language utterance; and
[27.b]	responsive to receipt of the second natural language utterance, generate a response to the second natural language utterance based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance.
[28.pre]	The system of claim 16, wherein the first non-voice input is received via a non-voice input device comprising a touch screen device, and wherein the one or more physical processors are further caused to:
[28.a]	receive information indicating a location on the touch screen device that received the first non-voice input.

U.S. Patent No. 10,297,249 (the “’249 patent”) claims “[a] computer-implemented method of facilitating natural language system responses.” EX1001, 20:28-29. The claims recite basing the responses “on one or more prior multi-modal device interactions” and comparing sound from multiple voice inputs. *Id.*, 20:30-31, 20:43-47. These features are conventional and this Petition demonstrates that the prior art renders each claim of the ’249 patent obvious. *See, e.g., id.*, 20:28-64; *see also* EX1002, ¶¶1-252.

I. OVERVIEW OF THE ’249 PATENT

A. Alleged Invention

The ’249 patent is purportedly directed to a device that “understand[s] free form human utterances.” EX1001, 2:13-17. An utterance may be part of a multimodal input. *Id.*, 7:38-40. For example, the “device may include a combination of a microphone and a touch-screen device,” where an input to the device includes “an utterance...relating to a portion of a display on the touch-screen device that the user is touching.” *Id.*, 7:38-44. The device may receive the utterance using multiple microphones. *Id.*, 11:7-9. In this case, the device “may collate inputs and cooperatively filter out sound by comparing a speech signal from the various microphones.” *Id.*, 11:9-11; EX1002, ¶¶32-33.

The device understands the utterance by using “a context determination process.” *Id.*, 3:61-63, 11:23-26. It also uses “short-term shared knowledge...to

proactively build and evaluate interaction with a user as a conversation progresses, as well as long-term shared knowledge to proactively build and evaluate interaction with the user over time.” *Id.*, 16:16-21. The device then “formulate[s a] personalized and conversational response” to the user. *Id.*, 16:63-66, 17:41-46. The response “may include a verbal...reply played through an output device.” *Id.*, 18:9-12; EX1002, ¶¶34-36.

B. Priority Date

The ’249 patent was filed as Application No. 14/691,445 (the “’445 application”) on April 20, 2015 and claims priority through a chain of applications to Application No. 11/580,926, filed October 16, 2006. EX1001; EX1002, ¶31. This Petition treats October 16, 2006 as the priority date for the ’249 patent.¹ *Id.*

II. IDENTIFICATION OF CHALLENGE

A. Statutory Grounds

Petitioner requests *inter partes* review and cancellation of claims 1-28 on the following §103 grounds. *See also* EX1002, ¶¶48-252.

Ground	Claims	Prior Art
1	1, 6, 9, 12-16, 21, and 25-28	Kennewick, Oviatt, Fujimoto

¹ Petitioner reserves the right to challenge the ’249 patent’s priority date in this or other proceedings.

Ground	Claims	Prior Art
2	2-5 and 17-20	Kennewick, Oviatt, Fujimoto, Shimomura
3	7, 8, and 22	Kennewick, Oviatt, Fujimoto, Matsuda
4	10, 11, 23, and 24	Kennewick, Oviatt, Fujimoto, Cooper

B. Prior Art²

1. *Kennewick (EX1004)*

Kennewick is a U.S. patent application published September 30, 2004. EX1004. Kennewick is prior art because it published more than one year before the '249 patent's priority date. EX1002, ¶62.

Kennewick describes a system that allows “users to ask natural language speech questions” and “present[s] results in a natural manner.” EX1004, ¶2. The system is “a complete speech-based information query, retrieval, presentation and command environment” that “makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands.” *Id.*, ¶11. Kennewick's system receives input using a **speech unit**, “processes the input..., and responds to the user with a natural language speech response.” *Id.*, ¶¶ 12, 14; EX1002, ¶¶63, 65.

² Pre-AIA 35 U.S.C. §102 applies to the '249 patent.

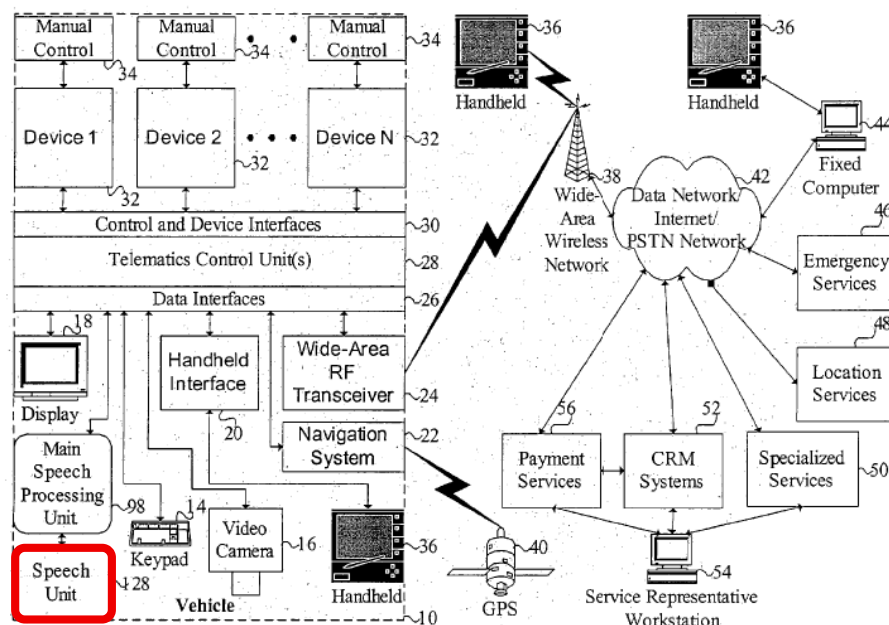


Figure 2. Second Embodiment System Block Diagram

EX1004, Figure 2, ¶108. “The speech unit 128 includes one or more microphones...to receive the utterances of the user.” *Id.*, ¶121; EX1002, ¶¶64-65.

2. Oviatt (EX1005)

Oviatt is a printed publication titled “Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions” by Sharon Oviatt et al. EX1005. Oviatt was published in the journal *Human-Computer Interaction* in 2000, in volume 15, pp. 263-322. *Id.*, 3. The declaration of Rachel J. Watters (EX1006) demonstrates that Oviatt was publicly accessible before the priority date of the ’249 patent. EX1006, 2-3. Oviatt is therefore prior art to the ’249 patent. EX1002, ¶¶66, 69-71.

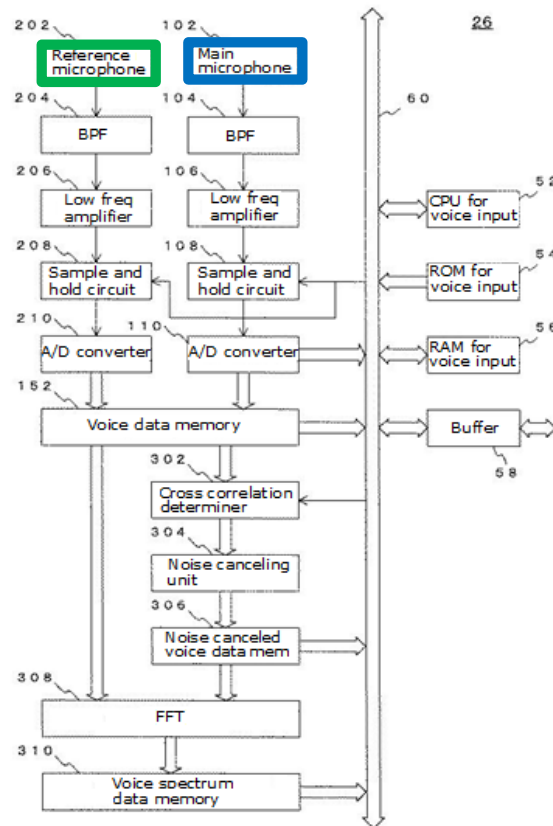
Oviatt “summarize[s] the emerging architectural approaches for interpreting speech and pen-based gestural input.” EX1005, 5. It “describe[s] a diverse collection of state-of the-art multimodal systems that process users’ spoken and gestural input.” *Id.* One such system is a Portable Voice Assistant (PVA), which “is a pen-voice multimodal interface.” *Id.*, 38. “The PVA can interpret simultaneous pen-voice input so, for example, the user can say ‘Show me that part’ while pointing at a screw on the display.” *Id.*, 39; EX1002, ¶¶67-68.

3. *Fujimoto (EX1007, EX1008)*

Fujimoto is a Japanese patent application published May 20, 2004. EX1007. Exhibit 1008 is a certified English translation of Fujimoto. Fujimoto is prior art because its publication date predates the priority date of the ’249 patent. EX1002, ¶72.

Fujimoto discloses “a voice matching device that compares voice signals of registered speakers with those of a speaker to be matched.” EX1007, ¶9. The voice signals are received by a “voice input device...equipped with a **main microphone** 102 and a **reference microphone** 202.” *Id.*, ¶42 (annotations added).³

³ All annotations and emphases are added unless otherwise noted.



Id., Fig. 4. The signal from the **reference microphone** 202 is “used to correct the main voice signals supplied by the **main microphone** 102.” *Id.*, ¶44; EX1002, ¶¶73-74.

4. *Shimomura (EX1014)*

Shimomura is a U.S. patent application published September 13, 2001. EX1014. Shimomura is prior art because it published before the '249 patent's priority date. EX1002, ¶75.

Shimomura discloses a robot “for holding a conversation with a user.” EX1014, ¶¶9, 18. Shimomura notes that when a person has a conversation, “it is not common for them to continue to discuss only one topic” because “[i]n general,

people change the topic at an appropriate point.” *Id.*, ¶95. Accordingly, Shimomura’s robot “has a function for changing the topic at an appropriate circumstance when having a conversation with a user” (*id.*, ¶96) “so that the conversation with the user continues naturally and fluently” (*id.*, ¶111). EX1002, ¶¶76-78.

5. Matsuda (EX1015, EX1016)

Matsuda is a Japanese patent application published November 25, 2004. EX1015. Exhibit 1016 is a certified English translation of Matsuda. Matsuda is prior art because its publication date predates the priority date of the ’249 patent. EX1002, ¶79.

Matsuda describes a conversation system using a “subject person number” to track the individual currently being discussed in a conversation. EX1015, ¶26. The number is reset if “at least a prescribed amount of time (for example, 10 minutes) has elapsed without a name of that person being mentioned.” *Id.*, ¶¶26, 38. Matsuda explains that this enables “appropriate statements corresponding to the conversation having moved away from the subject of that person.” *Id.*, ¶100; EX1002, ¶80.

6. Cooper (EX1017)

Cooper is a U.S. patent issued June 29, 2004. EX1017. Cooper is prior art because its issue date predates the priority date of the ’249 patent. EX1002, ¶81.

Cooper describes a “virtual assistant [that] adapts to the user in many different ways based on the input the virtual assistant receives.” EX1017, 2:56-64, Abstract. Cooper determines the user’s manner of speech and adapts the assistant’s output to match the user. For example, if the user uses polite words, the assistant will also use polite words. *Id.*, 43:55-61. EX1002, ¶82.

III. LEVEL OF ORDINARY SKILL IN THE ART

A person of ordinary skill in the art at the priority date of the ’249 patent (“POSITA”) 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, ¶¶23-26.

IV. CLAIM CONSTRUCTION

For the purposes of this Petition, Petitioner applies plain-and-ordinary meaning to all terms.⁴ EX1002, ¶¶27-30.

⁴ Petitioner reserves the right to offer different claim constructions in other forums.

V. GROUND 1: KENNEWICK, OVIATT, AND FUJIMOTO

A. Claim 1

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

Kennewick in combination with Oviatt discloses the preamble. EX1002, ¶¶84-103.

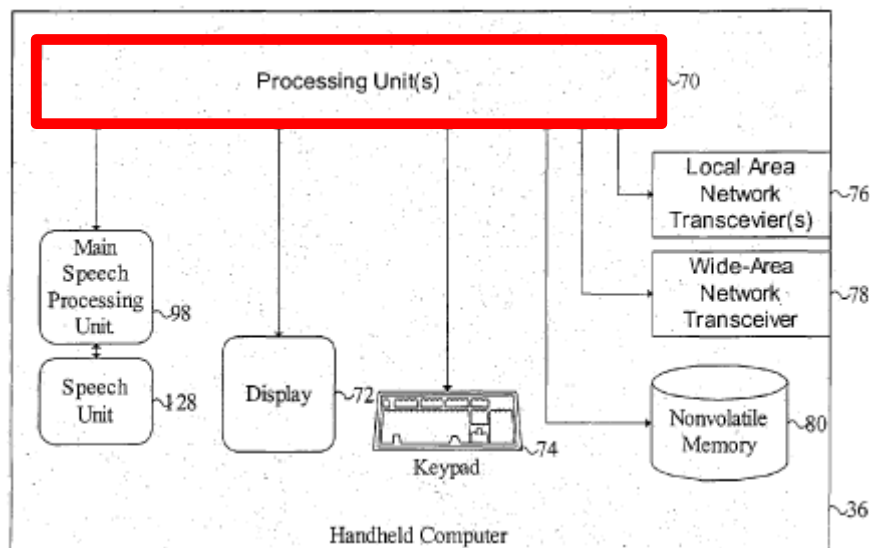
Kennewick describes an “integrated environment allowing mobile users to ask natural language speech questions or give natural language commands in a wide range of domains.” EX1004, ¶2. The environment “responds to the user with a natural language speech response.” *Id.*, ¶¶14, 23, 24, 27, 28, 118, 155-56, claim 28. The environment is provided by “systems and methods” (*id.*, ¶11) and therefore performs a *“method of facilitating natural language system responses.”* EX1002, ¶¶84-85.

Kennewick also discloses the method *“facilitating natural language system responses using short-term knowledge.”* The ’249 patent describes *“short-term knowledge”* as knowledge that “may accumulate during a single conversation, where input received during a single conversation may be retained.” EX1001, 5:6-8, 14:5-6, 16:15-19, 16:50-53, 17:17-20. EX1002, ¶86. Kennewick’s method likewise accumulates *“short-term knowledge”* during a dialog (a conversation) with a user.

The method adds “tags...to each utterance [in the dialog]...to tie that utterance to the correct...dialog” and user to build a dialog history for use in subsequent processing. EX1004, ¶¶155-56. The dialog history is “*short-term knowledge*” because it reflects details about a user’s utterances during the conversation. EX1002, ¶¶87-88; *see* EX1001, 16:50-53.

The method uses the “*short-term knowledge*” to *facilitat[e] natural language system responses*” to the user. For example, the method uses “probabilities or possibilities” values based on the dialog history to score possible contexts for an utterance. EX1004, ¶¶157, 162. The method also uses the dialog history to “score the relevance of the results” of a user query in order to facilitate a response to the user. *Id.*, ¶¶170, 171, 175; EX1002, ¶89.

Kennewick’s method is “*computer-implemented...by a computer system.*” The method may be implemented on numerous “computer alternatives” that Kennewick collectively refers to as “a computer.” EX1004, ¶13. The computer may be a “fixed computer” or “handheld device.” *Id.* Kennewick’s Figure 3 illustrates a possible embodiment of the handheld device.



EX1004, Figure 3, ¶¶90. These embodiments of the computer are each a “computer system” because they are formed of various computing components working together. *See id.*, ¶¶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, ¶¶90-91.

The “*computer system...includes one or more physical processors.*” Figure 3, shown above, illustrates that the computer system has “**processing units**” which “include one or more central processing units.” *Id.*, ¶¶111, 115, 101, 122, 107-09. Kennewick’s computer system thus “*includes one or more physical processors.*” EX1002, ¶92.

The processors “*execut[e] one or more computer program instructions which, when executed, perform the method.*” 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.” *Id.*, ¶13. The software is formed of computer program instructions which the physical processors execute to perform the method. EX1002, ¶93; *see* EX1004, ¶¶17, 107.

Kennewick also discloses that the computer system is a “*multi-modal device*.” The computer system includes a speech unit, “displays 18 including touch screen displays for text, graphics and video, keypads 14 for data input..., and a pointing device or stylus.” *Id.*, ¶97.

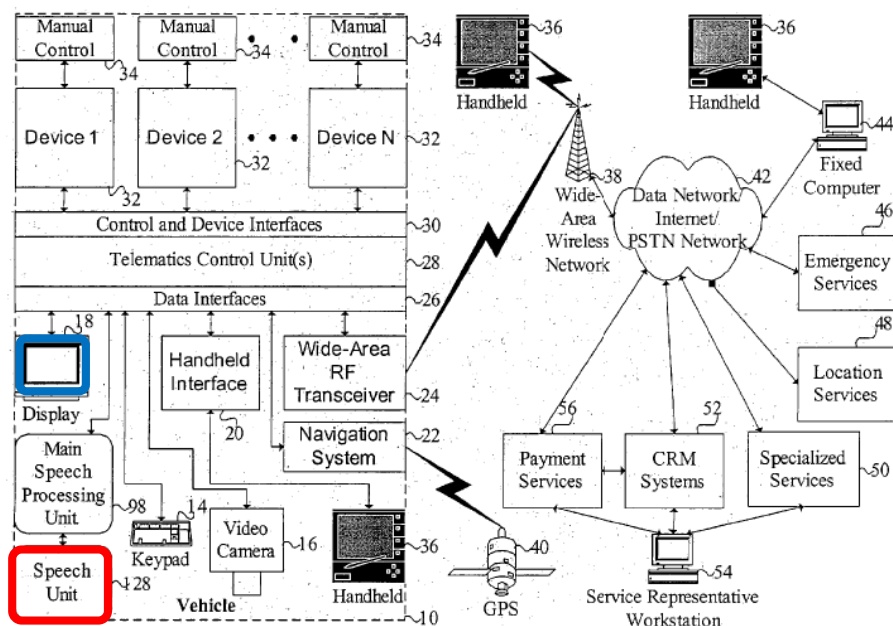
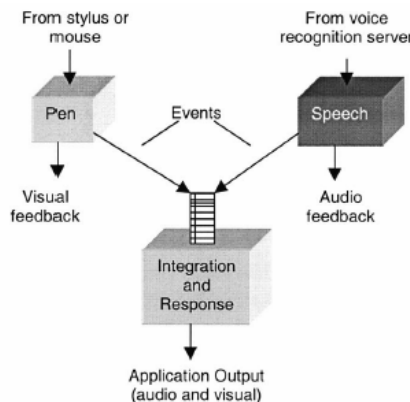


Figure 2. Second Embodiment System Block Diagram

Id., Figure 2. The computer system is a “***multi-modal device***” because it accepts different modes of input, including speech, touch screen, and keypad input. EX1002, ¶95. But Kennewick does not explicitly disclose that short-term

knowledge used by the system is “***generated based on one or more prior multi-modal device interactions.***” EX1002, ¶96.

This aspect is taught by Oviatt, which discloses a portable voice assistant (“PVA”), a “***multi-modal device***” that “can interpret simultaneous pen-voice input.” EX1005, 39. For example, the user can say “show me that part” “while pointing at a screw on the display.” *Id.* “[T]he PVA time stamps both speech and pen input events and integrates them to form a...description of the user’s request....” *Id.*



Id., 40. The generated events are “***prior multi-modal device interactions***” because the events describe multimodal interactions that occurred earlier. EX1002, ¶¶97-98.

A POSITA would have found it obvious to modify Kennewick to support multimodal inputs as taught by Oviatt, and to generate short-term knowledge from these inputs, thereby producing “***short-term knowledge generated based on one or more prior multi-modal device interactions.***” Kennewick and Oviatt are analogous references; both relate to human-computer interactions involving a speech interface.

See EX1004, ¶2; EX1005, 6. A POSITA would have been motivated to make this combination because interfaces involving multimodal input “are particularly effective for supporting mobile tasks, such as communications and personal navigation.” EX1005, 10; EX1002, ¶99.

For example, a POSITA would have been motivated to combine Kennewick and Oviatt in order to provide navigation services that leverage both voice and touch-screen input. Kennewick discloses that the computer system with multimodal input devices can be in a **vehicle** having a **navigation system** for personal navigation. EX1004, ¶¶15, 61, 96, 97.

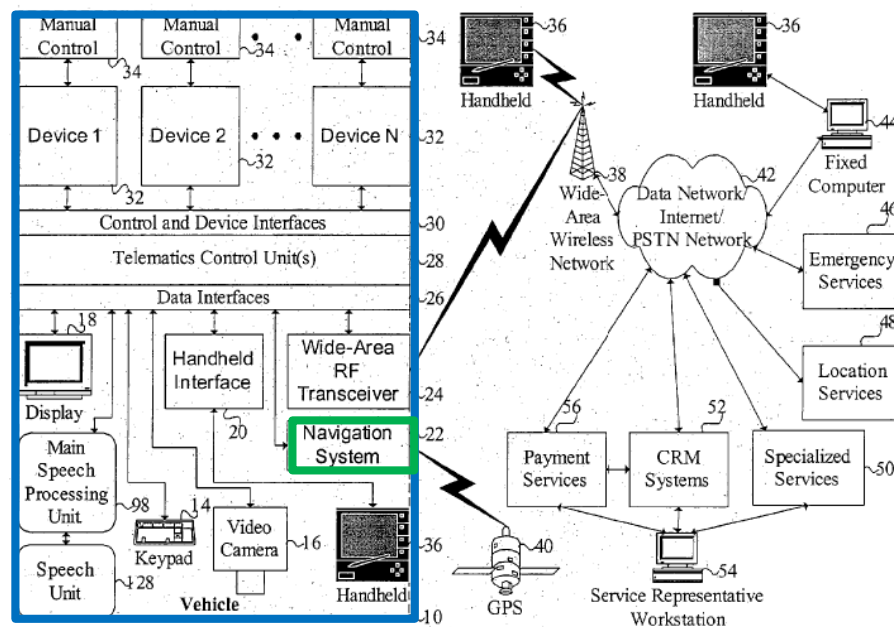


Figure 2. Second Embodiment System Block Diagram

EX1004, Figure 2. Oviatt discloses that the PVA “can interpret simultaneous pen-voice input” (EX1005, 39) and that “[f]uture field applications of this [PVA] system could include...in-car navigation systems.” EX1005, 40; EX1002, ¶100.

A POSITA would have found it obvious to incorporate Oviatt’s “*prior multi-modal device interactions*” into Kennewick’s short-term knowledge. This combination allows the user to use voice in combination with Kennewick’s existing touch-screen and stylus to control Kennewick’s navigation system. For example, the user could say “give me directions to here,” where the user indicates the place corresponding to “here” by pointing to a location on a map presented on the touch screen display. A POSITA would have been motivated to make this combination because “[a] pen stylus is a natural, precise, and effective mode of interaction for spatially oriented tasks.” EX1005, 29; EX1002, ¶101.

This is a combination of prior art elements according to known methods to yield predictable results. It is also the application of a known technique – Oviatt’s multimodal interactions involving a touch screen– to a known device and method – Kennewick’s computer having a touch screen – ready for improvement. Kennewick discloses a natural language speech interface and describes computer interfaces using “point and click techniques and touch screen displays” as conventional. EX1004, ¶¶3-4. Oviatt discloses prior art techniques involving multimodal device interactions including speech and touch-screen interactions. A POSITA would have

found it obvious to use Oviatt's techniques within Kennewick's system, would have found the results predictable, and would have reasonably expected the results to be successful, because Oviatt explains how to process multimodal inputs. EX1002, ¶¶102-103.

[1.a] “receiving, by the computer system during a first conversation, a first voice input via a first input device, the first voice input comprising a first natural language utterance;”

Kennewick discloses limitation [1.a]. EX1002, ¶¶104-107.

Kennewick's “**computer system**” includes a **speech unit**. EX1004, ¶¶12, 107.

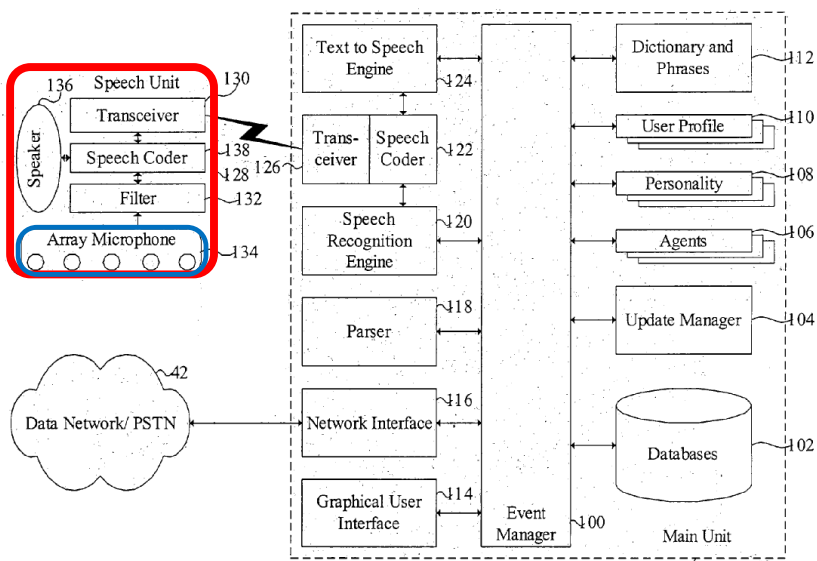


Figure 5. Speech Processing System Block Diagram

EX1004, Figure 5. The **speech unit** “**receiv[es] a first voice input via a first input device.**” The speech unit includes an **array microphone** that receives voice inputs. *Id.*, ¶¶38, 121, 149. A first microphone of the array is a “**first input device**” and “**receives a first voice input**” from a user. EX1002, ¶¶104-105. The “**first voice**

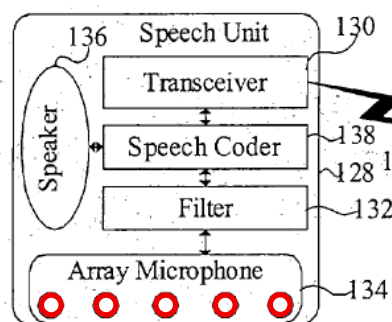
input” *“compris[es] a first natural language utterance”* because it is a natural language utterance provided by the user. EX1004, ¶¶2, 73, 101, 121; EX1002, ¶106; *see* EX1001, 2:26-28 (“an ‘utterance’ may be words, syllables, phonemes, or any other audible sound made by a human being”).

Kennewick also discloses that the “*computer system*” receives the “*first voice input*” “*during a first conversation.*” Kennewick discloses that the user has a “dialog with the system.” EX1004, ¶156. A dialog is a conversation. Thus, a natural language utterance received by “*the computer system*” during a dialog with the user is a “*first voice input comprising a first natural language utterance*” received “*during a first conversation.*” EX1002, ¶107.

[1.b] “receiving, by the computer system, a second voice input comprising the first natural language utterance via a second input device;”

Kennewick discloses limitation [1.b]. EX1002, ¶¶108-111.

Kennewick’s array microphone includes **multiple separate microphones** for receiving natural language utterances.



EX1004, Figure 5 (excerpt), ¶¶38, 121, 149, 151, 152; *see* EX1011, 530 (“[m]icrophone arrays use multiple microphones”); EX1002, ¶¶108-109. A second microphone of the array is a “***second input device***” because it is a second device that receives a voice input from a user. The voice input is “***a second voice input***” because it is distinct from the first voice input received by the first input device of limitation [1.a]. Therefore, Kennewick’s “***computer system***,” receives, “***via [the] second input device***,” “***a second voice input***.” EX1002, ¶110.

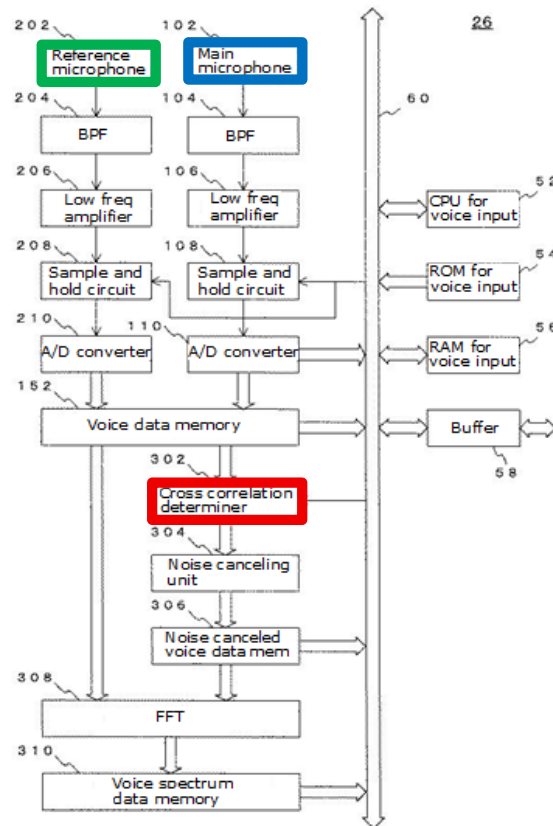
The second voice input “***compris[es] the first natural language utterance***.” The first and second input devices are in close proximity, e.g., adjacent microphones in the array. Therefore, both microphones receive the same user utterance. *See, e.g.*, EX1004, ¶¶121 (“[t]he speech unit 128 includes one or more microphones, for example array microphone 134, to receive the utterances of the user”), 147, 149; EX1011, 530-531. Kennewick’s “***computer system***” thus receives “***a second voice input comprising the first natural language utterance via [the] second input device***.” EX1002, ¶111.

[1.c] “comparing, by the computer system, the first voice input with the second voice input;”

Kennewick in combination with Fujimoto discloses limitation [1.c]. EX1002, ¶¶112-117.

Kennewick discloses “*the computer system*” receiving “*the first voice input*” and “*the second voice input*” as discussed for limitation [1.b]. Fujimoto discloses “*comparing the first voice input with the second voice input.*” EX1002, ¶112.

Fujimoto discloses a “voice input device 26...equipped with a **main microphone 102** and a **reference microphone 202.**” EX1008, ¶42. The “**main microphone 102** receives main voice signals.” *Id.*, ¶43. The **reference microphone 202**, in contrast, “receives voice signals...that are used to correct the main voice signals supplied by the **main microphone 102.**” *Id.*, ¶44; EX1002, ¶113. A **cross correlation determiner 302** compares the signal from the **main microphone** with the signal from the **reference microphone**:



EX1008, Figure 4. In particular, the **cross correlation determiner** “compares the main voice data with the reference data, and as a result of this comparison, extracts the reference data that is most similar to the main voice data from the reference data.” *Id.*, ¶83. It then “outputs the main voice data and the extracted reference voice data.” *Id.*, ¶63; EX1002, ¶114.

A POSITA would have found it obvious to use Fujimoto’s comparison within Kennewick’s computer system to “*compar[e]...the first voice input with the second voice input.*” Both references relate to receiving user utterances. *See, e.g.*, EX1004, ¶2; EX1008, ¶9. Both references also disclose receiving voice inputs via multiple

input devices. EX1002, ¶115; *see, e.g.*, EX1004, ¶121; EX1008, ¶77. Given these similarities, a POSITA would have been motivated to modify Kennewick's computer system to compare the first and second voice inputs as part of a process to reduce noise in order to improve speech recognition. Kennewick's computer system uses multiple microphones "to provide better...noise elimination than can be achieved with a single microphone." EX1004, ¶147. Kennewick also discloses that "noise elimination may be a key aspect of the invention" and "[t]he accurate recognition and parsing of the user's speech requires the best possible signal to noise ratio at the input to the speech recognition engine 120." *Id.*, ¶146. Fujimoto discloses comparing first and second voice inputs as part of a noise reduction process. EX1008, ¶83; EX1002, ¶116.

This is a combination of prior art elements according to known methods to yield the predictable result of noise reduction. It is also the application of a known technique (cross-correlation) as taught in Fujimoto to a known device (Kennewick's computer system) ready for improvement to yield the predictable result of extracting data for use in noise reduction. A POSITA would have had a reasonable expectation of success because comparing two signals is well-understood, and Kennewick states that noise reduction can be performed with its multiple microphones and leads to more "accurate recognition and parsing of the user's speech." EX1004, ¶146; EX1002, ¶117.

[1.d] “filtering, by the computer system, sound from the first voice input and the second voice input based on the comparison;”

Kennewick in combination with Fujimoto discloses limitation [1.d]. EX1002, ¶¶118-126.

Kennewick in combination with Fujimoto discloses “*the comparison*” per limitation [1.c]. Fujimoto additionally discloses “*filtering...sound from the first voice input and the second voice input based on the comparison.*” EX1002, ¶118.

Fujimoto’s cross-correlation determiner “outputs the main voice data and the extracted reference voice data.” EX1008, ¶63. This output is “*sound from the first voice input*” (the main voice data) “*and the second voice input*” (the extracted reference voice data). The output is received by Fujimoto’s noise canceling unit 304, which then “corrects the main voice data and removes noise.” *Id.*, ¶64; EX1002, ¶¶119-120.

A POSITA would have understood that Fujimoto’s noise-removal process involves “*filtering.*” A “filter” is “hardware or software that selectively passes certain elements of a signal and eliminates or minimizes others.” EX1020, 3. Fujimoto’s noise canceling unit 304, which receives two signals (the main voice data and the extracted reference voice data) and outputs a signal that “corrects the main voice data and removes noise,” is “filtering” the input signals by removing noise (i.e., by eliminating or minimizing noise). EX1002, ¶¶121-122; *see* EX1008, ¶¶64, 84. The “*filtering*” is “*based on the comparison*” because the noise canceling unit

operates on the main voice data and the extracted reference voice data output by the cross correlation determiner that performed the comparison. EX1008, ¶¶63-64; EX1002, ¶123. Therefore, Fujimoto discloses “*filtering...sound from the first voice input and the second voice input based on the comparison.*” EX1002, ¶124.

A POSITA would have found it obvious to perform Fujimoto’s noise canceling within Kennewick’s computer system to “*filter[]...sound from the first voice input and the second voice input based on the comparison*” for the same reasons as for limitation [1.c]. A POSITA would have been motivated to modify Kennewick’s computer system to perform Fujimoto’s noise canceling in order to reduce noise and improve speech recognition. EX1004, ¶¶146-48; EX1008, ¶¶64, 84; EX1002, ¶125.

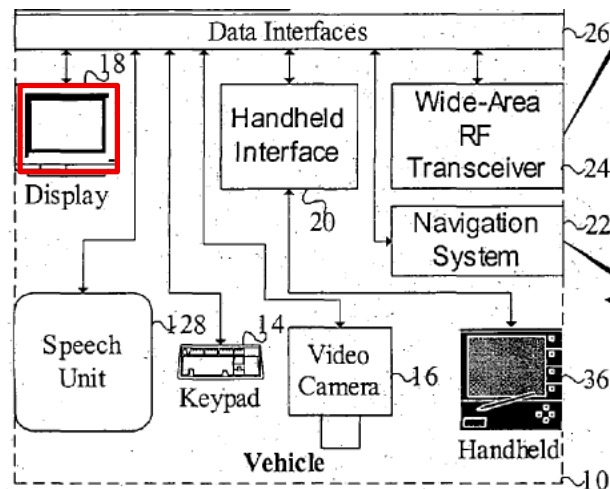
This is a combination of prior art elements according to known methods to yield the predictable result of noise reduction. It is also the application of a known technique (noise reduction) as taught in Fujimoto to a known device (Kennewick’s computer system) ready for improvement to yield the predictable result reducing noise in a signal. A POSITA would have had a reasonable expectation of success because noise reduction is well-understood and leads to more “accurate recognition and parsing of the user’s speech.” EX1004, ¶146; EX1002, ¶126.

[1.e] “obtaining, by the computer system during the first conversation, a user interface state related to one or more non-voice inputs associated with the first

voice input, the one or more non-voice inputs comprising at least a first non-voice input;

Kennewick in combination with Oviatt discloses limitation [1.e]. EX1002, ¶¶127-136.

Kennewick discloses that the computer system includes input devices for “***obtaining***” “***one or more non-voice inputs.***” EX1004, ¶97. For example, the computer system may include “**touch screen displays** [18] for text, graphics and video, keypads 14 for data input..., and a pointing device or stylus.” *Id.*

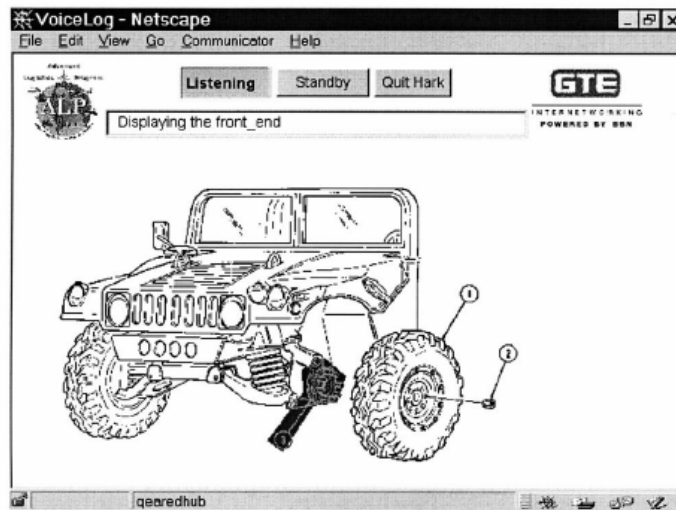


EX1004, Figure 1 (excerpt), ¶¶112, 116. These types of non-voice input devices are conventional. *Id.*, ¶3; EX1002, ¶128.

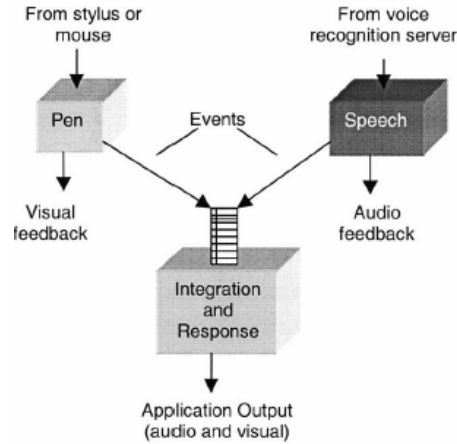
Kennewick also discloses that a non-voice input device has “***a user interface***” with which the user can interact. A “graphical user interface 114 can be used as a...complement to the speech interface.” EX1004, ¶125. “For example, the graphical user interface can be used to view and interact with graphical or tabular

information in a manner more easily digested by the user...[and] [t]he graphical user interface can show system state and history in a more concise manner than the speech interface.” *Id.*; EX1002, ¶129.

However, Kennewick does not explicitly disclose limitation [1.e]. This is taught by Oviatt’s PVA, which has “*a user interface*” illustrating images the user can select using a pen and touch screen:



EX1005, 39 (Figure 13). The PVA “can interpret simultaneous pen-voice input so, for example, the user can say ‘Show me that part’ while pointing at a screw on the display.” *Id.*, 39 “[T]he PVA time stamps both speech and pen input events and integrates them to form a frame-based description of the user’s request, as illustrated in Figure 14.” *Id.*



Id., 40 (Figure 14); EX1002, ¶¶130-131.

Oviatt here discloses “*obtaining...a user interface state related to one or more non-voice inputs...comprising at least a first non-voice input.*” The “*first non-voice input*” is the user “pointing at a screw on the display.” EX1005, 39. The PVA “*obtain[s]...a user interface state related to one or more non-voice inputs...comprising at least a first non-voice input*” by receiving an event indicating when the user pointed to (touched) the display and to where the user pointed on the user interface presented by the display (e.g., the part pointed to by the user). *Id.*, 39-40. This event indicates the state of the user interface when the user points to it, and the “*user interface state*” is therefore “*related to one or more non-voice inputs...comprising at least a first-non-voice input.*” EX1002, ¶132.

Oviatt also discloses that the “*non-voice input[] is associated with the first voice input.*” An integrator thread associates the voice and non-voice inputs based on time. Thus, when the user says “‘Show me that part’ while pointing at a screw

on the display” “the integrator simply would look for a pen event that occurred closest in time to the spoken deictic event ‘that.’” EX1005; EX1002, ¶133.

Given the complementary disclosures of Kennewick and Oviatt, a POSITA would have found it obvious to combine the references so Kennewick’s computer system “*obtain[s]...during the first conversation, a user interface state related to one or more non-voice inputs associated with the first voice input, the one or more non-voice inputs comprising at least a first non-voice input during the first conversation.*” For example, a POSITA would have found it obvious to combine the references in the navigation context, where a user, “*during the first conversation*” with Kennewick’s computer system utters “give me directions to here” while indicating the location corresponding to “here” by touching a location on the touch screen display. Kennewick discloses that the user converses with a computer system having a touch screen display and “navigation system components.” EX1004, ¶97. Oviatt discloses “*obtaining...a user interface state related to [a] non-voice input[] associated with the first voice input,*” and suggests using this disclosure in a navigation system. EX1005, 40; EX1002, ¶134. A POSITA would have been motivated to combine the references in this manner because “[i]nterfaces involving spoken or pen-based input, as well as the combination of both, are particularly effective for supporting mobile tasks, such as

communications and personal navigation.” EX1005, 10; *see also id.*, 8; EX1002, ¶135.

This is a combination of prior art elements according to known methods to yield predictable results. Kennewick discloses using voice inputs in a computer system having a conventional touch screen. EX1004, ¶3. Oviatt discloses a known way to effectively combine both voice and non-voice inputs on a computer system having a touch screen. EX1005, 38-40. The combination is also the application of a known technique (Oviatt’s combination of voice and non-voice inputs) to a known device ready for improvement (Kennewick’s computer system with voice input and a touch screen) to yield predictable results, and use of a known technique improve a similar device (computer with voice and touch screen inputs) in the same way. A POSITA would have found the combination predictable because these are well-understood technologies, and would have reasonably expected the combination to be successful for the same reason. EX1002, ¶136.

[1.f] “generating, by the computer system, the short-term knowledge based on at least the first voice input and the first non-voice input;”

Kennewick combined with Oviatt discloses limitation [1.f]. EX1002, ¶¶137-141.

Kennewick’s dialog history is short-term knowledge because it reflects details about a user’s utterances during a conversation. EX1004, ¶¶23, 25, 29, 31, 37, 155, 157, 160. Kennewick’s computer system generates the short-term knowledge by

systematically tagging each utterance of the dialog. *Id.*, ¶155. Kennewick thus discloses the “*computer system*” “*generating...the short-term knowledge based on at least the first voice input.*” EX1002, ¶138.

Oviatt discloses “*generating...the short-term knowledge based on at least first voice input and the non-voice input*” because it receives voice and non-voice inputs and generates short-term knowledge based on these inputs. Oviatt’s PVA “can interpret simultaneous pen-voice input [events]...and integrates them to form a frame-based description of the user’s request.” EX1005, 39. These events and the description are “*short-term knowledge based on at least first voice input and the first non-voice input*” because they include a history of the user’s voice and non-voice inputs with the PVA built as the user’s conversation progresses. EX1002, ¶¶139-140; *see* EX1001, 16:50-53.

A POSITA would have found it obvious to combine Kennewick and Oviatt so “*the computer system*” “*generat[es] the short-term knowledge based on at least the first voice input and the first non-voice input.*” For example, a POSITA would have found it obvious for Kennewick’s computer system to incorporate Oviatt’s first non-voice input into the dialog history by tagging it along with the first voice input to generate the short-term knowledge because Kennewick teaches tagging voice inputs and Oviatt teaches using a similar approach for voice and non-voice inputs. *See, e.g.*, EX1004, ¶155; EX1005, 62. A POSITA also would have found it obvious

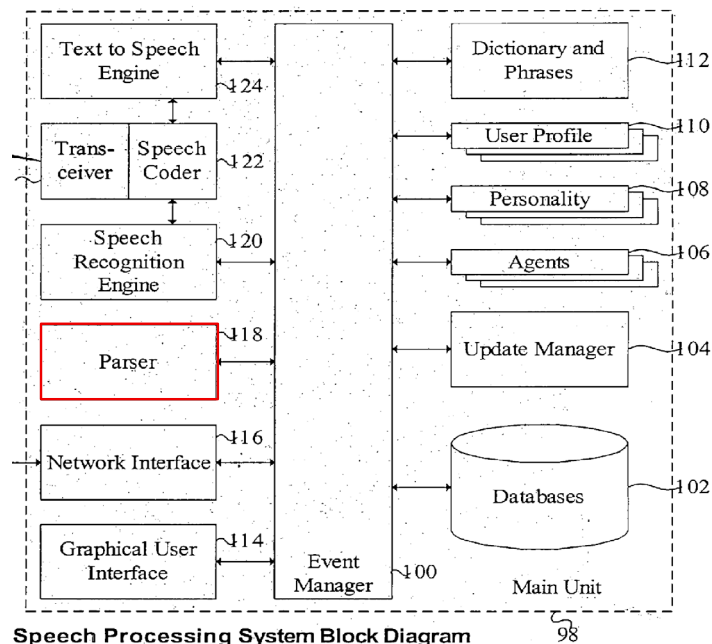
to store the history information Oviatt's PVA generates based on the inputs as part of Kennewick's short-term knowledge. A POSITA would have been motivated to combine the references in these ways to allow Kennewick's system to use multimodal inputs, which are "are particularly effective for supporting mobile tasks, such as communications and personal navigation" (EX1005, 10) in addition to the other motivations provided for the preamble and limitation [1.e]. EX1002, ¶141.

[1.g] "determining, by the computer system, based on the short-term knowledge, a first context for the first natural language utterance;"

Kennewick discloses limitation [1.g]. EX1002, ¶¶142-144.

Kennewick discloses the "*computer system*" "*determining...a first context for the first natural language utterance*" "*based on the*" dialog history, which is "*short-term knowledge.*" EX1004, ¶¶23, 25, 29, 31, 37, 155, 157, 160.

Kennewick's speech processing unit includes a **parser** that determines the context for an utterance based on the dialog history. EX1004, ¶160 ("[t]he parser may determine a context for an utterance by applying prior probabilities or fuzzy possibilities to...[the] dialog history").



EX1004, Figure 5 (excerpt). Kennewick also explains that the computer system uses the dialog history to determine context for an utterance because “some questions can only be interpreted in the context of previous questions” in the dialog. *Id.*, ¶7. Accordingly, Kennewick discloses and renders obvious the “*computer system*” “*determining...based on the short-term knowledge, a first context for the first natural language utterance*” because the parser uses the dialog history, which is short-term knowledge, to determine the (first) context for the first utterance. EX1002, ¶¶142-144.

[1.h] “*determining, by the computer system, based on the first context, an interpretation of the first natural language utterance; and*”

Kennewick discloses limitation [1.h]. EX1002, ¶¶145-148.

Kennewick's parser interprets natural language utterances received from the user. EX1004, ¶¶28, 41, 143, 155; EX1002, ¶146. It determines the interpretations based on the contexts of the utterances. EX1004, ¶7. In particular, the parser uses recognized words and phrases, context, and other criteria to determine an appropriate context-based agent to be invoked for a user utterance forming a question or command. *Id.*, ¶¶160, 162, 165. The invoked agent works with the parser to interpret the utterance based on the context. For example, if the user asks "what about flight one hundred and twenty too?," the parser may determine that the context is "travel planning" and invoke a travel planning domain agent specific to this context, where "[t]he parser and domain agent use flight information in the database and network information along with context to determine the most plausible interpretation among; flight 100 and flight 20 also, flight 100 and flight 22, flight 122, etc." *Id.*, ¶163; EX1002, ¶147. Thus, Kennewick discloses "***determining, by the computer system, based on the first context, an interpretation of the first natural language utterance.***" EX1002, ¶148.

[1.i] "generating, by the computer system, based on the interpretation of the first natural language utterance, a first response to the first natural language utterance."

Kennewick discloses limitation [1.i]. EX1002, ¶¶149-152.

The invoked agent works with the parser to interpret the utterance based on the context. EX1004, ¶¶162, 163. The agent also generates a response to the

utterance. *Id.*, ¶17 (“[a]gents are executables that receive, process and respond to user questions, queries and commands”), 32, 124, 178; EX1002, ¶150. For example, the agent can generate a response and send it to a text-to-speech engine that speaks the response to the user. EX1004, ¶¶124, 178; EX1002, ¶151.

A POSITA would have understood and found obvious that the response addresses the interpreted utterance. A POSITA likewise would have understood and found obvious that the response is a first response to the first utterance because it may be the only response or a first of multiple responses to the first utterance. Accordingly, Kennewick discloses and renders obvious the “*computer system*” “*generating..., based on the interpretation of the first natural language utterance, a first response to the first natural language utterance.*” EX1002, ¶152.

B. Claim 6

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

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. Kennewick discloses the additional limitation of claim 6. EX1002, ¶¶153-154.

Kennewick’s parser interprets “*one or more recognized words of the first natural language utterance.*” EX1004, ¶¶28, 41, 143, 155, 160 (“[o]nce the words and phrases have been recognized by the speech recognition engine 120, the tokens and user identification may be passed to the parser 118”). Therefore, when the parser

“determin[es] the interpretation of the first natural language utterance” per limitation [1.h] it is *“determining, based on the first context, an interpretation of one or more recognized words of the first natural language utterance.”* EX1002, ¶154.

C. Claim 9

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

Kennewick-Oviatt-Fujimoto teaches the method of claim 1.

[9.a] “accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received during conversations occurring prior to the first conversation, wherein the first context is determined based on the short-term knowledge and the long-term knowledge.”

Kennewick discloses limitation [9.a]. EX1002, ¶156-159.

Kennewick discloses the *“computer system” “accumulating...long-term knowledge.”* The ’249 patent describes long-term knowledge as “generally...user-centric, rather than session-based, where inputs may be accumulated over time to build user, environmental, cognitive, historical, or other long-term knowledge models.” EX1001, 5:26-30; EX1002, ¶156. Kennewick’s computer system *“accumulat[es]”* information from past or historical conversations in a user profile, by “creat[ing], stor[ing] and us[ing] extensive personal profile information for each user.” EX1004, ¶31. The user profile includes, among other things, a “history of questions asked,” “session histories,” a “history of commands and queries,” and a

“history of [individual] interactions with the system.” *Id.*, ¶¶31, 143. Therefore, the information within the user profile, which is user-centric and accumulated over multiple sessions, is “*accumulat[ed]...long-term knowledge.*” EX1002, ¶157.

The “*long-term knowledge is accumulated...prior to the first conversation.*”

A POSITA would have understood that a user may have multiple conversations with Kennewick’s computer system, including conversations before the first conversation, and that the computer system receives natural language utterances from the user during these conversations. EX1004, ¶156. Accordingly, the computer system “*receive[s]*” “*natural language utterances...during conversations occurring prior to the first conversation*” and “*the long-term knowledge is accumulated based on*” these utterances, such as the “history of questions asked,” “session histories,” “history of commands and queries,” and “history of [individual] interactions with the system.” *Id.*, ¶¶31, 143; EX1002, ¶158.

Kennewick also discloses that “*the first context is determined based on the short-term knowledge and the long-term knowledge.*” Kennewick’s “parser may determine a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents.” EX1004, ¶160. The dialog history is short-term knowledge and the user profile includes long-term knowledge. Therefore, “*the first context is determined based on the short-term knowledge and the long-term knowledge.*” EX1002, ¶159.

D. Claim 12

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

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. EX1002, ¶160.

Kennewick discloses that ***“the second input device is included with the first input device in a single device.”*** The speech unit in which the first and second input devices are included is a single device. EX1004, ¶12 (“The speech unit may be incorporated into a vehicle computer device or system, or may be a separate device.”). The speech unit may also be part of another single device, such as a handheld device. *Id.*, ¶96; *see also id.*, ¶¶107, 111, 115; EX1002, ¶162. Moreover, including the second input with the first input device in a single device would have been obvious because Kennewick discloses that the speech unit can operate in different configurations. EX1004, ¶109; EX1002, ¶163.

E. Claim 13

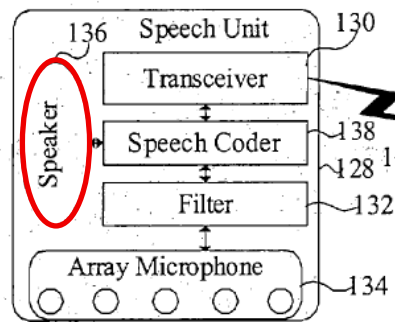
“The method of claim 1, wherein generating the first response comprises generating, by the computer system, a non-voice response in a graphical user interface and/or a voice response via a speaker.”

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. EX1002, ¶164.

Kennewick alone and in combination with Oviatt discloses the additional limitation of claim 13. *Id.*, ¶¶165-167.

Kennewick discloses that ***“generating the first response comprises generating, by the computer system...a voice response via a speaker.”*** The first

response generated by the computer system may be “a natural language speech response,” which is a “*a voice response*.” EX1004, ¶14. The voice response is “*via a speaker*” because it is “output by the **speaker 136**” of the speech unit. *Id.*, ¶124.



EX1004, Figure 5 (excerpt); *see also id.*, ¶121; EX1002, ¶165.

Kennewick also discloses that “*generating the first response comprises generating, by the computer system, a non-voice response in a graphical user interface*.” Kennewick’s computer system has a “graphical user interface 114 [that] can be used as a substitute or complement to the speech interface.” EX1004, ¶125. It would have been obvious for Kennewick’s computer system to generate the first response as “*a non-voice response in a graphical user interface*” because “[s]ome information, in formats such as video, pictures and graphics, may be best presented in a displayed format.” *Id.*, ¶182; EX1002, ¶166.

In addition, the computer system “*generating a non-voice response in a graphical user interface*” is obvious in view of Kennewick and Oviatt. Oviatt discloses that the output of multimodal systems “typically is presented via a

graphical, auditory (e.g., for telephony), or multimedia display.” EX1005, 15. A POSITA would have been motivated to combine Kennewick and Oviatt such that Kennewick’s computer system generates and presents a non-voice output as taught by Oviatt because “[s]ome information, in formats such as video, pictures and graphics, may be best presented in a displayed format.” EX1004, ¶182; EX1002, ¶167.

F. Claim 14

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

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. EX1002, ¶168. Kennewick teaches preamble [14.pre.] EX1002, ¶¶169-170.

Kennewick further teaches wherein “***the first natural language utterance relates to a request.***” Kennewick teaches that “[t]he system...can process and respond to questions, queries and commands.” EX1004, ¶27. A command is a “request” because a command requests that the computer system perform an action. *Id.*, ¶¶123, 166. For example, a user may utter a request that the computer system tune a radio by uttering the command “tune in my favorite radio station.” *Id.*, ¶27; EX1002, ¶169.

Kennewick also discloses “***wherein...the first response is intended to be responsive to the request.***” The first response generated by the computer system is

“intended to be responsive to request” because it is a reply to the request. For example, if the computer system does not understand the user’s request, “the system can ask a question of the user to verify the question or command is correctly understood,” such as “[d]id I understand that you want such-and-such.” EX1004, ¶161; *see also id.*, ¶¶32, 33, 177; EX1002, ¶170.

[14.a] “receiving, by the computer system, a second natural language utterance in response to the first response, wherein the second natural language utterance comprises a clarification related to the interpretation of the first natural language utterance; and”

Kennewick teaches limitation [14.a]. EX1002, ¶171.

Kennewick’s computer system provides a first response that asks the user for ***“a clarification related to the interpretation of the first natural language utterance.”*** For example, if the user’s command is ambiguous, the computer system can ask the user “one or more questions to resolve the ambiguity.” EX1004, ¶¶32, 33, 161, 177. A POSITA would have understood and found obvious that the user provides, and the computer system receives, ***“a second natural language utterance in response”*** to the question from the system, and the ***“utterance comprises a clarification related to the interpretation of the first natural language utterance.”*** *See, e.g., id.*, ¶¶33 (“the user can rephrase the original question, perhaps adding additional information to remove ambiguity”), 161 (same); EX1002, ¶171.

[14.b] “responsive to receipt of the second natural language utterance, generating, by the computer system, a response to the second natural language utterance

based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance.”

Kennewick discloses limitation [14.b]. EX1002, ¶172.

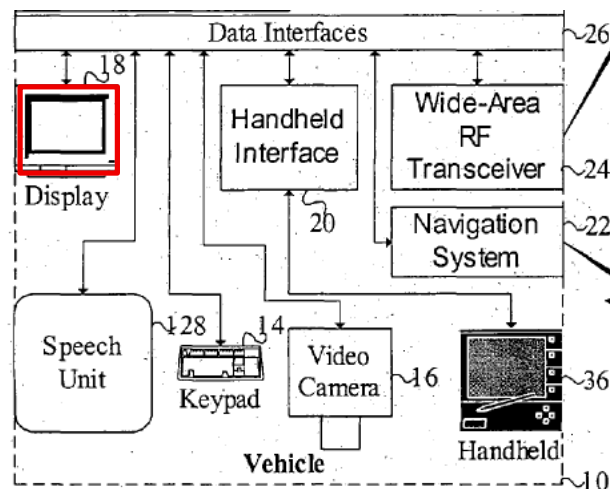
The second utterance the computer system receives from the user clarifies a command the computer system misunderstood or found ambiguous. *Id.* The computer system responds to receipt of the user’s clarification (“***responsive to receipt of the second natural language utterance***”) by generating a response to the clarified command (“***generating...a response to the second natural language utterance***”). EX1004, ¶32 (“The user can then be asked one or more questions to resolve the ambiguity. Additional queries may then be made before an adequate response is made.”). The system’s response is based on both the interpretation of the first utterance and the user’s clarification in the second utterance (“***a response...based on the interpretation of the first natural language utterance and the clarification related to the interpretation of the first natural language utterance***”). *Id.*, 33 (“If the user confirms that the question is correct then the system may proceed to produce a response.”), 161; EX1002, ¶172.

G. Claim 15

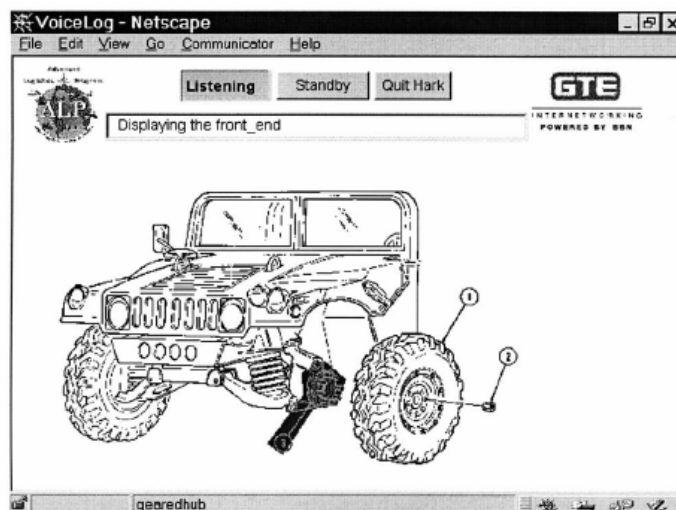
[15.pre] “The method of claim 1, wherein the first non-voice input is received via a non-voice input device comprising a touch screen device, the method further comprising:”

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. EX1002, ¶173.

Kennewick and Oviatt disclose “*receiv[ing]*” “*the first non-voice input*” for the reasons provided for limitation [1.e]. These references also disclose “*a non-voice input device comprising a touch screen device.*” Kennewick’s computer system has a touch screen **display**, which is a “*non-voice input device comprising a touch screen device.*” EX1004, ¶¶97.



EX1004, Figure 1 (excerpt), ¶¶112, 116. Oviatt’s PVA also has a touch screen display.



EX1005, 39 (Figure 13). EX1002, ¶174.

Oviatt discloses that “*the first non-voice input is received via a non-voice input device comprising a touch screen device.*” Oviatt’s PVA “can interpret simultaneous pen-voice input so, for example, the user can say ‘Show me that part’ while pointing at a screw on the display.” EX1005, 39. The user “pointing at a screw on the display” is the first non-voice input and is received via the display. EX1002, ¶175.

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

Oviatt discloses and renders obvious limitation [15.a]. EX1002, ¶¶176-177.

Oviatt’s PVA “can interpret simultaneous pen-voice input so, for example, the user can say ‘Show me that part’ while pointing at a screw on the display.” EX1005, 39. In this situation, the PVA interprets the pen input by “*receiving...information indicating a location on the touch screen device*” to which the user is pointing. *See, e.g., id.*, 8 (“pen input to maps and other graphic displays can easily and efficiently convey spatial information about precise points, lines, and areas”), 10 (“pen input conveys rich and precise graphical information”). The touch screen device for which the location is received is same device “that received the first non-voice input.” Thus, the input is “**receiv[ed]...by the computer system...that received the first non-voice input.**” EX1002, ¶177.

H. Claim 16

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

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

Claim 16 is unpatentable for the same reasons as claim 1. Preamble [16.pre] and limitation [16.a] are substantively identical to preamble [1.pre]. Preamble [16.pre] and limitation [16.a] recite “[a] system...comprising...one or more physical processors programmed with one or more computer program instructions which, when executed, cause the one or more physical processors to” perform recited operations. Preamble [1.pre] similarly recites “[a] computer-implemented method...implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method” comprising recited steps. The “computer system” implementing the method of claim 1 is equivalent to the “system” of claim 16 and obvious for the same reasons. EX1002, ¶178.

The remaining limitations of claim 16 are substantively identical to the corresponding limitations of claim 1 and obvious for the same reasons. EX1002, ¶179.

I. Claims 21 and 25-28

Dependent claims 21 and 25-28 are unpatentable for the same reasons as the respective corresponding claim indicated in the following table:

<u>Claim Number</u>	<u>Corresponding Claim Number</u>
21	6
25	12
26	13
27	14
28	15

Each of claims 21 and 25-28 is substantively identical to the corresponding claim, and contains only minor differences reflecting that independent claim 16 and any intervening claims recite a system. EX1002, ¶¶180-181.

VI. GROUND 2: KENNEWICK, OVIATT, FUJIMOTO, AND SHIMOMURA

A. Claim 2

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

Kennewick-Oviatt-Fujimoto discloses the method of claim 1. EX1002, ¶182.

[2.a] “receiving, at the computer system, a second natural language utterance via the first input device;”

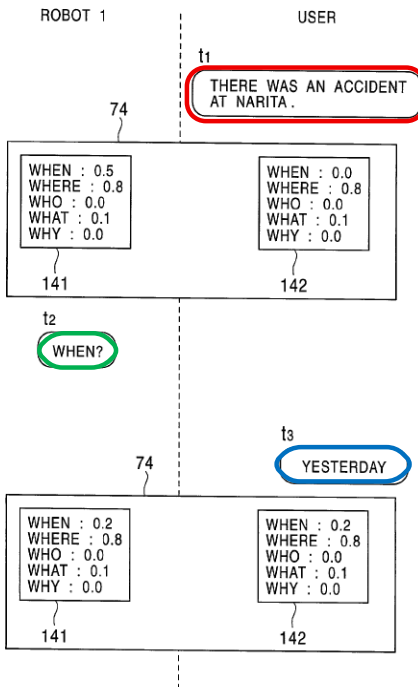
Kennewick alone and in combination with Shimomura discloses limitation [2.a]. EX1002, ¶¶183-187.

Kennewick discloses “*the computer system*” “*receiving...a second natural language utterance*” as discussed for limitation [14.a]. *See, e.g.*, EX1004, ¶¶32, 33, 161, 177. Kennewick’s computer system receives the utterance “*via*” a first microphone of the array microphone, which is “*the first input device.*” *Id.*, ¶121. Accordingly, Kennewick discloses the “*computer system*” “*receiving...a second natural language utterance via the first input device.*” EX1002, ¶184.

Kennewick in combination with Shimomura also discloses limitation [2.a]. Shimomura discloses a robot with “a conversation processing apparatus for holding a conversation with a user.” EX1014, ¶¶9, 17. The robot includes a microphone that “collects ambient speech (sounds) including the speech of a user.” *Id.*, ¶56. The robot thus “*receiv[es]*” utterances from the user via the microphone, which is a “*first input device.*” EX1002, ¶185.

Shimomura discloses that a conversation between the user and robot includes multiple utterances. For example, Shimomura describes the user initiating a conversation with the robot by saying “[t]here was an accident at Narita.” The robot responds with “when?” and the user answers “[y]esterday.” EX1014, ¶¶122-25.

FIG. 12



Id., Fig. 12. This conversation includes first and second natural language utterances from the user: 1) “there was an accident at Narita”; and 2) “yesterday.” These utterances are captured by the robot’s microphone (a first input device) since the microphone “collects ambient speech (sounds) including the speech of a user.” *Id.*, ¶56; EX1002, ¶186.

A POSITA would have found it obvious to combine Kennewick and Shimomura to support dialogs in which the “*computer system*” “*receiv[es]...a second natural language utterance via the first input device.*” Both Kennewick and Shimomura relate to natural language interactions with users. See EX1004, ¶2; EX1014, ¶2. This combination would support Kennewick’s goal of “allow[ing] a

person to make natural language speech-based queries,” especially given that “that most natural language queries and commands are incomplete in their definition” and may require follow-up utterances. EX1004, ¶¶5, 7-8. This is a combination of prior art elements according to known methods to yield predictable results, and a POSITA would have reasonably expected the combination to be successful because Kennewick discloses how the computer system processes utterances. EX1002, ¶187.

[2.b] “determining, by the computer system, based on the short-term knowledge, whether the second natural language utterance corresponds to the first context;”

Kennewick in combination with Shimomura discloses limitation [2.b]. EX1002, ¶¶188-195.

Kennewick’s computer system determines a context for each utterance it receives. EX1004, ¶¶28 (“[f]or each user utterance including a question or query or set of questions or queries, the system may perform multiple steps possibly including...determining the...context”), 41, 160. It determines the context for an utterance “**based on short-term knowledge.**” EX1004, ¶160. Accordingly, a POSITA would have found it obvious that Kennewick’s computer system determines the context for the second natural language utterance “**based on the short-term knowledge.**” EX1002, ¶189.

Kennewick’s computer system also determines the context for the second natural language utterance based on the first context for the first natural language

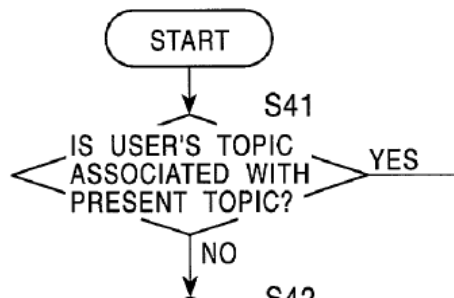
utterance. The parser “use[s] a scoring system to determine the most[] likely context or domain for a user’s question or command.” EX1004, ¶161. “The score can be determined from weighting a number of factors including, the...previous context.” *Id.*; see also *id.*, ¶¶32, 160; EX1002, ¶190.

Given this disclosure, a POSITA would have understood and found obvious that the “*computer system*” “*determin[es] based on the short-term knowledge, whether the second natural language utterance corresponds to the first context*” by using the “scoring system to determine the most[] likely context or domain for a user’s question or command.” EX1004, ¶161. A POSITA would have understood that the context for the second utterance may be different than the context of the first utterance. See *id.*, ¶¶167, 177 (“real-time selection of the correct agent allows the invention to dynamically switch contexts”). A POSITA also would have known that “some questions can only be interpreted in the context of previous questions.” *Id.*, ¶7. Therefore, a POSITA would have found it obvious for the scoring system to determine whether the context for the second utterance changed from the context of the first utterance, and would have understood that this determination involves determining “*whether the second natural language utterance corresponds to the first context.*” EX1002, ¶191.

To the extent Kennewick does not explicitly disclose “*determining...whether the second natural language utterance corresponds to the first context,*” it would

have been obvious based on Shimomura. Shimomura's robot talks with the user about a topic. The robot detects when a user's utterance changes the topic. EX1014, ¶¶184, 195.

FIG. 25



Id., Fig. 25 (excerpt); EX1002, ¶192.

The robot detects topic changes by comparing the topic introduced by the user with the current topic. EX1014, ¶184 (“[i]n step S41, the topic manager 74 of the robot 1 determines whether the topic introduced by the user is associated with the present topic stored in the present topic memory 77”). It performs the comparison using a scoring system. *Id.*, ¶¶142, 143, 184, 185. Accordingly, a POSITA would have understood Shimomura to disclose determining whether a second natural language utterance corresponds to a topic of a first natural language utterance. EX1002, ¶193.

A POSITA would have found it obvious in view of Shimomura for Kennewick's “*computer system*” to “*determin[e]...whether the second natural language utterance corresponds to the first context*” when determining the context

for the second utterance. Kennewick's parser "use[s] a scoring system to determine the most[] likely context or domain for a user's question or command" where "[t]he score can be determined from...the...previous context." EX1004, ¶161. Shimomura discloses an analogous scoring system for detecting topic changes that determines whether a second utterance corresponds to a topic of a first utterance. EX1014, ¶185. Therefore, a POSITA would have found it obvious for Kennewick's computer system to use a process for contexts analogous to the process for topics disclosed in Shimomura to detect context changes by "*determining...whether the second natural language utterance corresponds to the first context.*" EX1002, ¶194.

A POSITA would have been motivated to make this combination because "a person asking a question or giving a command typically relies heavily on context" and "some questions can only be interpreted in the context of previous questions" EX1004, ¶¶5, 7. Accordingly, a POSITA would have been motivated to use Shimomura's technique within Kennewick's computer system in order "to hold natural conversations" by determining whether the second utterance corresponds to the (first) context of the first utterance. EX1014, ¶201. This is a combination of prior art elements according to known methods to yield predictable results and use of a known technique to improve a similar device in the same way. A POSITA

would have had a reasonable expectation of success because the technology at issue is conventional and well understood. EX1002, ¶195.

[2.c] “determining, by the computer system, responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge;”

Kennewick discloses and renders obvious limitation [2.c]. EX1002, ¶¶196-199.

Kennewick discloses that the computer system determines a context for each received utterance based on the short-term knowledge as discussed for limitations [1.g] and [2.b]. EX1004, ¶¶28, 41, 160, 162; EX1002, ¶197. A POSITA would have understood and found obvious that the computer system determines a different context for the second natural language utterance if it does not correspond to the first context. The computer system uses the “scoring system to determine the most[] likely context or domain for a user’s question or command,” with the outcome that the context is different from the first context. EX1004, ¶161; *see id.*, ¶¶167, 177 (“real-time selection of the correct agent allows the invention to dynamically switch contexts”). Hence, Kennewick discloses and renders obvious the “***computer system***” “***determining...responsive to the second natural language utterance not corresponding to the first context, a second context for the second natural language utterance based on the short-term knowledge.***” EX1002, ¶198.

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

Kennewick’s computer system determines an interpretation of a natural language utterance based on the determined context as discussed for limitation [1.h]. See, e.g., EX1004, ¶¶160, 162, 165. For limitation [2.d], a POSITA would have understood that the “*computer system*” “*determin[es], based on the second context, an interpretation of the second natural language utterance*” in the same manner it interprets the first utterance based on the first context. EX1002, ¶199.

[2.e] “generating, by the computer system, based on the interpretation of the second natural language utterance, a response to the second natural language utterance.”

Kennewick’s computer system generates a response to a natural language utterance based on the interpretation of the utterance as discussed for limitation [1.i]. See, e.g., EX1004, ¶¶14 (“the system...receives spoken natural language queries, commands and/or other utterances from a user,...processes the input...and responds to the user with a natural language speech response”), 17, 32, 124, 178. For limitation [2.e], a POSITA would have understood that the “*computer system*” “*generat[es]...based on the interpretation of the second natural language utterance, a response to the second natural language utterance*” in the same manner it generates the first response to the first natural language utterance based on the interpretation as discussed for limitation [1.i]. EX1002, ¶200.

B. Claim 3

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

Kennewick-Oviatt-Fujimoto discloses the method of claim 1. EX1002, ¶201.

[3.a] “tracking, by the computer system, contexts identified for multiple consecutive natural language user utterances to generate a context stack, the context stack including the contexts identified for the multiple consecutive natural language user utterances;”

Kennewick in combination with Shimomura discloses limitation [3.a]. EX1002, ¶¶202-208.

Kennewick alone and in combination with Shimomura discloses “*the computer system*” receiving “*multiple consecutive natural language user utterances.*” As discussed for limitations [14.a] and [2.a], Kennewick discloses examples of dialogs between the computer system and user having multiple natural language utterances. *See, e.g., id.*, ¶¶32, 33, 161, 177. These dialogs include “*multiple consecutive natural language user utterances.*” For example, Kennewick describes the user rephrasing a question the computer system finds ambiguous. The initial question is a first natural language user utterance and the rephrased question is a second, consecutive natural language user utterance. *See id.*, ¶33; EX1002, ¶203.

Shimomura explicitly discloses conversations including “*multiple consecutive natural language user utterances.*” For example, Shimomura describes the user initiating a conversation with the robot by saying “[t]here was an

accident at Narita.” The robot responds with “when?” and the user answers “[y]esterday.” EX1014, ¶¶122-25. Here, the user’s utterances “[t]here was an accident at Narita” and “[y]esterday” are “*multiple consecutive natural language user utterances.*” EX1002, ¶204.

Kennewick discloses “*identify[ing] contexts...for...multiple ...natural language user utterances*” as discussed for limitations [1.g], [2.b], and [2.c]. The computer system identifies a context for each utterance it receives. EX1004, ¶¶28, 41, 160. A POSITA would have found it obvious from Kennewick’s disclosure to “*identify contexts...for...multiple consecutive natural language user utterances*” since this would beneficially allow the user to converse with the computer system by detecting potential context changes. EX1002, ¶205; *see* EX1004, ¶¶160, 162, 165, 167.

Kennewick also discloses that the computer system uses “*a context stack.*” *Id.*, ¶170 (“part of the dialog history is maintained in a context stack”); *see also id.*, ¶160. A stack is a conventional data structure for tracking and storing information. *See, e.g.*, EX1011, 642 (describing storing data “in a last-in first-out (LIFO) stack”). A POSITA therefore would have understood and found it obvious that Kennewick’s “*computer system*” “*generate[s] a context stack...including the contexts identified for the multiple consecutive natural language user utterances*” and uses it to

“track[]...contexts identified for multiple consecutive natural language user utterances.” EX1002, ¶206.

To the extent not disclosed in Kennewick, Shimomura discloses a data structure storing and tracking information about a conversation, including information about previous conversation topics:

The conversation history memory 75 accumulates the content of conversation or information extracted from conversation. The conversation history memory 75 also stores data used to examine topics which were brought up prior to the present topic, which is stored in the present topic memory 77, and to control the change of topic.

EX1014, ¶¶83, 176. A POSITA would have understood from this disclosure that Shimomura’s robot updates the topics stored in the conversation history memory when it changes topics based on user utterances. *Id.*, ¶82; EX1002, ¶207.

A POSITA would have found it obvious in view of Kennewick and Shimomura for Kennewick’s *“computer system”* to *“track[]...contexts identified for multiple consecutive natural language user utterances to generate a context stack...including the contexts identified for the multiple consecutive natural language user utterances.”* Kennewick explicitly discloses a context stack and may identify the context for each utterance, including for multiple consecutive natural language user utterances. *See, e.g.*, EX1004, ¶¶160, 170. Shimomura discloses storing analogous information about a conversation in a data structure for tracking

purposes. *See, e.g.*, EX1014, ¶83. A POSITA would have found it obvious from these disclosures to generate a context stack and store contexts in it for tracking purposes. A POSITA would have been motivated to use the context stack to provide an “environment allowing mobile users to ask natural language speech questions or give natural language commands in a wide range of domains” because “[t]he context of a question or command may determine the domain and thereby, the domain agent 156, if any, to be invoked” and the possible contexts are scored using information in the context stack. EX1004, ¶¶2, 160. This is a combination of prior art elements according to known methods to yield predictable results. EX1002, ¶208.

[3.b] “receiving, at the computer system, a second natural language utterance via the first input device;”

Kennewick alone and in combination with Shimomura discloses “*the computer system*” “*receiving...a second natural language utterance via the first input device*” as discussed for limitations [14.a] and [2.a]. EX1002, ¶209.

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

Kennewick in combination with Shimomura discloses limitation [3.c]. EX1002, ¶¶210-211. As discussed for limitation [2.b], Kennewick and Shimomura disclose the “*computer system*” determining the context for the second natural language utterance based on the short-term knowledge and determining whether the second natural language utterance corresponds to the first context for a first (prior)

utterance. Kennewick and Shimomura also disclose a context stack that includes contexts for prior utterances as discussed for limitation [3.a]. EX1002, ¶210.

Given these disclosures, a POSITA would have found it obvious for the “computer system” to “*determin[e]..., based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.*” A POSITA would have found it obvious that the computer system determines whether the second natural language utterance corresponds to the first context per limitation [2.b] by determining whether the second utterance corresponds to the first context (an individual one of the contexts), or another individual context, included in the context stack per limitation [3.a]. *See, e.g.*, EX1004, ¶¶160, 170 (“[t]he weight of each context for the scoring may be based on the relevance of one context to another and the age of the contexts”); *see also id.*, ¶¶83, 176. A POSITA would have been motivated to use the context stack this way because Kennewick’s computer system “determine[s] a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents” and the context stack stores contexts of prior utterances. *Id.*, ¶160; EX1002, ¶211.

C. Claim 4

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

Kennewick in combination with Shimomura discloses the method of claim 3, including “*determining whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack.*” EX1002, ¶212.

[4.a] “determining whether the second natural language utterance corresponds to a most recent context included in the context stack; and”

Limitation [4.a] is obvious in view of Kennewick and Shimomura. These references disclose including contexts in the context stack, and determining whether the second natural language utterance corresponds to an individual context included in the context stack as discussed for limitation [3.c]. EX1002, ¶213.

A POSITA would have found it obvious to “*determin[e] whether the second natural language utterance corresponds to a most recent context included in the context stack*” because a stack is a LIFO (last-in first-out) data structure. EX1011, 642. A POSITA would have understood that the top of the stack includes the last-in (most recent) context. In addition, Kennewick discloses weighing the influence of a context based on the age of the context. EX1004, ¶¶160, 170 (“[t]he weight of each context for the scoring may be based on the relevance of one context to another and the age of the contexts”); *see also id.*, ¶¶83, 176. Therefore, a POSITA would

have found it obvious for the computer system to determine whether the second natural language utterance corresponds to a context in the stack by determining whether the second utterance corresponds to the most recent (i.e., youngest) context included at the top of the stack. EX1002, ¶214.

[4.b] “responsive to the second natural language utterance not corresponding to the most recent context, determining whether the second natural language utterance corresponds to a second most recent context included in the context stack.”

Limitation [4.b] is obvious in view of Kennewick and Shimomura. Following on from limitation [4.a], a POSITA would have understood and found obvious that ***“responsive to the second natural language utterance not corresponding to the most recent context”*** at the top of the stack, the computer system would ***“determin[e] whether the second natural language utterance corresponds to a second most recent context included in the context stack,”*** which is the next context included in the context stack because the stack is a LIFO data structure. EX1004, ¶170 (“[t]he weight of each context for the scoring may be based on the relevance of one context to another and the age of the contexts”). EX1002, ¶215.

D. Claim 5

[5.pre] “The method of claim 3, further comprising:”

Kennewick in combination with Shimomura discloses the method of claim 3. EX1002, ¶216.

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

Kennewick and Shimomura disclose the computer system determining, based on the short-term knowledge, whether the second natural language utterance corresponds to one or more individual ones of the contexts included in the context stack as discussed for limitation [3.c]. Kennewick also discloses the computer system determining, responsive to the second natural language utterance not corresponding to a first context, a second context for the second natural language utterance based on the short-term knowledge as discussed for limitation [2.c]. EX1002, ¶216.

Given these disclosures, a POSITA would have found it obvious for the “computer system” to “*determin[e]...a second context for the second natural language utterance*” in the claimed manner. That is, if the second utterance does not correspond to contexts in the context stack, Kennewick’s computer system would obviously determine a second context for the second utterance. EX1004, ¶¶28, 41, 160. And it would determine the second context based on the short-term knowledge for the reasons provided for limitation [2.c]. *Id.* Kennewick accordingly discloses the “computer system” “*determining...a second context for the second natural language utterance based on the short-term knowledge*” “*responsive to the*

second natural language utterance not corresponding to one or more individual ones of the contexts included in the context stack.” EX1002, ¶218.

[5.b] “determining, by the computer system, based on the second context, an interpretation of the second natural language utterance.”

Kennewick discloses limitation [5.b]. As explained for limitation [1.h], Kennewick discloses the computer system determining an interpretation of the first natural language utterance based on the first context. Kennewick discloses the “computer system” “determining...an interpretation of the second natural language utterance” “based on the second context” for the same reasons. EX1002, ¶219.

E. Claims 17-20

Dependent claims 17-20 are unpatentable for the same reasons as the respective corresponding claim indicated in the following table:

<u>Claim Number</u>	<u>Corresponding Claim Number</u>
17	2
18	3
19	4
20	5

Each of claims 17-20 is substantively identical to the corresponding claim, and contains only minor differences to reflect that independent claim 16 and any intervening claims recite a system. EX1002, ¶¶220-221.

VII. GROUND 3: KENNEWICK, OVIATT, FUJIMOTO, AND MATSUDA

A. Claim 7

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

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. EX1002, ¶222.

[7.a] “expiring, by the computer system, one or more items of short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior to the first conversation.”

Kennewick alone and in combination with Matsuda discloses limitation [7.a]. EX1002, ¶¶223-231.

Kennewick discloses “*one or more items of short-term knowledge that are based on one or more natural language utterances received during*” a conversation as discussed for preamble [1.pre] and limitation [1.a]. EX1004, ¶155. Kennewick also discloses “*receiv[ing] one or more natural language utterances...during conversations occurring prior to the first conversation*” as discussed for limitation [9.a]. EX1002, ¶223.

It is obvious from Kennewick’s disclosure for the computer system to “*expir[e]...one or more items of short-term knowledge that are based on one or more natural language utterances received during conversations occurring prior*

to the first conversation.” As discussed for limitation [1.a], Kennewick’s computer system accumulates “short-term knowledge” reflecting details about a user’s utterances during a conversation. EX1004, ¶¶23, 25, 29, 31, 37, 155, 157, 160. Since a user can have multiple conversations with the computer system, the system accumulates short-term knowledge based on one or more natural language utterances received during prior conversations, just as it accumulates short-term knowledge during a current conversation. EX1002, ¶224.

In addition, Kennewick’s computer system obviously “*expir[es]*” short-term knowledge associated with prior conversations because the expired knowledge no longer reflects details about a user’s utterances during a current conversation. For example, the computer system may expire short-term knowledge from a prior conversation at the beginning the first conversation. EX1002, ¶225.

To the extent Kennewick does not explicitly disclose “*expiring, by the computer system, one or more items of short-term knowledge,*” Matsuda does. Matsuda discloses a conversation system that receives natural language utterances from a user. EX1016, ¶36. The system stores information identifying a current conversation subject in a memory, which enables a conversation “to progress naturally while returning appropriate responses in relation to the subject.” *Id.*, ¶12. EX1002, ¶226.

Matsuda uses a “subject person number” which “indicates a person that is the present subject of the conversation.” *Id.*, ¶26. Matsuda stores the subject person number in a working memory. *Id.*, ¶38. If the subject person number in the working memory is 0 there is no current subject person. *Id.*, ¶26. If the number is at least 1, then the person indicated by the number is the subject. *Id.* EX1002, ¶227.

Matsuda’s system tracks elapsed time since the subject person was last mentioned by the user. *Id.*, ¶¶27, 69. The system expires the subject person number if the elapsed time exceeds a specified amount. Specifically, “if a prescribed time 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. *Id.*, ¶38. This way, the system avoids referencing the wrong person in the conversation. *Id.*, ¶100. EX1002, ¶228.

The subject person number stored in the working memory is short-term knowledge because it identifies the “present subject of the conversation.” *Id.*, ¶26; *see* EX1001, 16:50-53 (describing “short-term knowledge” as including a “history of the conversation” built “as the conversation progresses”). Accordingly, Matsuda discloses “*expiring, by the computer system, one or more items of short-term knowledge;*” it expires the subject person number which is short-term knowledge. EX1002, ¶229.

A POSITA would have been motivated to use Matsuda's technique for expiring short-term knowledge within Kennewick's computer system. Kennewick and Matsuda are analogous art; both relate to voice-based interactions with computers. *See, e.g.*, EX1004, ¶2; EX1016, ¶10. Kennewick's computer system uses short-term knowledge and seeks to "achieve a natural environment for one or more users making queries or commands." EX1004, ¶11. A POSITA would have found it obvious to incorporate Matsuda's teachings into Kennewick because Matsuda discloses that expiring short-term knowledge "makes it possible to achieve a natural conversation." EX1016, ¶116. EX1002, ¶230.

For example, it would have been obvious to expire Kennewick's short-term knowledge about a prior conversation at the beginning of a new conversation, such as by clearing Kennewick's dialog history much like how Matsuda clears the subject person number stored in the working memory. Doing so would allow the new conversation "to progress naturally while returning appropriate responses" as taught by Matsuda because the new conversation would not use the dialog history from the prior conversation as short-term knowledge. EX1016, ¶12. This is a combination of prior art elements according to known methods to yield predictable results. A POSITA would have had a reasonable expectation that the combination would be successful because Matsuda teaches how to expire short-term knowledge. EX1002, ¶231.

B. Claim 8

[8.pre] “The method of claim 7, the method further comprising:”

Kennewick-Oviatt-Fujimoto-Matsuda teaches the method of claim 7.
EX1002, ¶232.

[8.a] “accumulating, by the computer system, long-term knowledge based on the one or more items of short-term knowledge that are expired.”

Kennewick and Matsuda disclose limitation [8.a]. EX1002, ¶¶233-235.

Kennewick discloses the “***computer system***” “***accumulating...long-term knowledge***” as discussed for limitation [9.a]. Kennewick and Matsuda disclose “***one or more items of short-term knowledge that are expired***” as discussed for limitation [7.a]. For example, the computer system may expire short-term knowledge from a prior conversation at the beginning of a new conversation, thereby producing “***one or more items of short-term knowledge that are expired.***” EX1002, ¶234.

Given this disclosure, it would have been obvious for Kennewick’s “computer system” to “***accumulat[e]...long-term knowledge based on the one or more items of short-term knowledge that are expired.***” Kennewick’s computer system “***accumulat[es]***” information from past or historical conversations in a user profile, by “creat[ing], stor[ing] and us[ing] extensive personal profile information for each user.” EX1004, ¶31. The user profile includes, among other things, a “history of questions asked,” “session histories,” a “history of commands and queries,” and a “history of [individual] interactions with the system.” *Id.*, ¶¶31, 143. A POSITA

therefore would have found it obvious to store the expired short-term knowledge as long-term knowledge within the user profile, e.g., as part of the “session histories” because Kennewick teaches that the user profile stores this type of information. *See id.*, ¶143. Hence, the “*accumulat[ed]...long-term knowledge [is] based on the one or more items of short-term knowledge that are expired.*” EX1002, ¶235.

C. Claim 22

[22.pre] “The system of claim 16, wherein the one or more physical processors are caused to:”

Preamble [22.pre] is unpatentable for the same reasons as preamble [16.pre] and limitation [16.a]. EX1002, ¶236.

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

Limitation [22.a] corresponds to limitation [7.a]. Limitation [22.a] recites “*expir[ing] one or more items of accumulated short-term knowledge*” while limitation [7.a] recites expiring “items of short-term knowledge.” Per preamble [1.pre], Kennewick’s computer system uses accumulated short-term knowledge to facilitate natural language system responses. *See, e.g.*, EX1004, ¶155. Therefore, the expired short-term knowledge is accumulated, and limitation [22.a] is obvious for the same reasons as limitation [7.a]. EX1002, ¶237.

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

Limitation [22.b] corresponds to limitations [8.a] and [9.a]. The first clause of limitation [22.b], “*accumulate long-term knowledge based on the one or more items of accumulated short-term knowledge that are expired,*” is nearly identical to limitation [8.a], except limitation [22.b] recites “*accumulated short-term knowledge*” while limitation [8.a] recites “short-term knowledge.” Kennewick discloses “*accumulated short-term knowledge*” per limitation [22.a]. Accordingly, the first clause of limitation [22.b] is obvious for the same reasons as limitations [8.a] and [22.a]. EX1002, ¶238.

The second clause of limitation [22.b] recites “*wherein the first context is determined based on short-term knowledge and the long-term knowledge.*” This clause is identical to the last clause of limitation [9.a] and is obvious for the same reasons. EX1002, ¶239.

VIII. GROUND 4: KENNEWICK, OVIATT, FUJIMOTO, AND COOPER

A. Claim 10

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

Kennewick-Oviatt-Fujimoto teaches the method of claim 1. EX1002, ¶240.

[10.a] “determining, by the computer system, based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the

first response is generated based on the determined manner and the interpretation of the first natural language utterance.”

Kennewick discloses the “***computer system***” receiving a first voice input comprising a first natural language utterance per limitation [1.a], generating short-term knowledge based on the first voice input per limitation [1.f], and “***generating***” “***the first response...based on...the interpretation of the first natural language utterance***” per limitation [1.i]. EX1002, ¶241. Kennewick does not disclose the computer system “***determining,...based on the short-term knowledge, a manner in which the first natural language utterance is spoken***” or that “***the first response is generated based on the determined manner.***” *Id.*

These aspects are taught by Cooper (EX1017). *Id.* Cooper describes a “virtual assistant [that] adapts to the user in many different ways based on the input the virtual assistant receives.” *Id.*, 2:56-64, Abstract. The input includes a natural language utterance from a user. *Id.*, Abstract, 6:35-37, 7:50-53, 46:58-59. EX1002, ¶242.

Cooper’s virtual assistant “***determin[es],...based on...short-term knowledge, a manner in which the first natural language utterance is spoken.***” The virtual assistant determines “whether words associated with polite discourse are included in input received from the user.” *Id.*, 43:55-59, 2:26-30. Cooper also explains that “the user information input into the virtual assistant could be information about the user’s emotion, which could be based on information about the user’s voice volume, word

choice and speech rate. Based on such information, the virtual assistant could automatically determine the user's emotional state, calm or angry, for example.” *Id.*, 43:62-67. Cooper thus “*determin[es],...a manner in which the first natural language utterance is spoken*” by determining whether the utterance includes words associated with polite discourse (*id.*, 43:55-59) and by determining whether the user spoke in a manner conveying an emotional state such as being calm or angry (*id.*, 43:62-67). Cooper determines the manner “*based on...short-term knowledge*” because Cooper uses information about natural language utterances received from the user during a conversation to make the determination. *See* EX1001, 5:6-7, 14:5-6, 16:15-19. EX1002, ¶243.

Cooper also discloses that “*the first response is generated based on the determined manner.*” If “words of polite discourse are included in the input received from the user, the virtual assistant could automatically enable the politeness setting.” EX1017, 43:59-61. The politeness setting “cause[s] the virtual assistant to include words or phrases associated with polite discourse in the output from the virtual assistant.” *Id.*, 43:49-51. The virtual assistant can similarly generate a response adapted to the user's emotional state. *Id.*, 44:1-4; EX1002, ¶244. Cooper's disclosure is consistent with disclosure in the '249 patent which describes creating a response “by adapting to a user's manner of speaking, framing responses appropriately, and having natural variation and/or personality (e.g., by varying tone,

pace, timing, inflection, word use, jargon, and other variables in a verbal or audible response).” EX1001, 6:24-28, 18:35-40; *see also id.*, 6:29-32; EX1002, ¶245.

A POSITA would have found it obvious to combine Cooper with Kennewick such that Kennewick’s “*computer system*” “*determin[es]...based on the short-term knowledge, a manner in which the first natural language utterance is spoken, wherein the first response is generated based on the determined manner and the interpretation of the first natural language utterance.*” Kennewick and Cooper are analogous art; both relate to voice-based computer interactions. *See, e.g.*, EX1004, ¶2; EX1017, 2:36-39. Kennewick’s computer system generates short-term knowledge from the dialog with the user. EX1004, ¶¶23, 25, 29, 31, 37, 155, 157, 160. Cooper discloses using short-term knowledge including the words of the dialog to determine the manner in which the user is speaking. EX1017, 43:55-59, 44:1-4. Accordingly, a POSITA would have found it obvious to determine the manner in which the first natural language utterance is spoken based on Kennewick’s short-term knowledge, because short-term knowledge includes information about the dialog in which the user made the utterance. *See, e.g.*, EX1004, ¶155. A POSITA also would have found it obvious to store other types of short-term knowledge taught by Cooper within Kennewick’s short-term knowledge to determine the user’s manner of speaking. EX1002, ¶246.

A POSITA would have been motivated to provide the user with natural interactions with the computer system, and would have been motivated to use Cooper's teachings to further this goal by adapting the computer system's interactions, including the generated first response, to match the manner of the user. *See, e.g.*, EX1004, ¶11; EX1017, 2:25-32, 2:44-45. This is a combination of prior art elements according to known methods to yield the predictable result of matching the response from the computer system to the user's manner of speaking. It is also use of the known technique for improving human-computer interactions disclosed in Cooper to improve human-computer interactions disclosed in Kennewick in the same way. A POSITA would have had a reasonable expectation of success because the technologies involved are conventional. EX1002, ¶247.

B. Claim 11

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

Kennewick-Oviatt-Fujimoto-Cooper teaches the method of claim 10. As discussed for limitation [10.a], Cooper discloses determining “***the manner in which the first natural language utterance is spoken***” by using “information about the user's emotion, which could be based on information about the user's voice volume, word choice and speech rate.” EX1017, 43:62-65; EX1002, ¶248.

The user's speech rate includes information about the pace and timing of the user's speech because these are factors that influence the rate of speech. In addition, the user's voice volume includes information about the tone and inflection of the user's speech because a user can convey tone and inflection by varying the volume of their speech. Accordingly, Cooper discloses that the manner “*includes at least one of a tone, a pace, an inflection, or a timing.*” EX1002, ¶249.

C. Claims 23 and 24

Claim 23 is unpatentable for the same reasons as claim 10. Claim 24 is unpatentable for the same reasons as claim 11. These claims are substantively identical to the corresponding claims, and contain only minor differences to reflect that independent claim 16 and any intervening claims recite a system. EX1002, ¶250-251.

IX. CONCLUSION

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

X. STANDING

Petitioner certifies that Petitioner is not barred or estopped from requesting this *inter partes* review and the '249 patent is IPR eligible.

XI. 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 '249 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 is concurrently filing a petition for *inter partes* review of U.S. Patent No. 10,755,699, which is related to the '249 patent.

Petitioner is aware of the following additional related matters involving the '249 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 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

Petition for *Inter Partes* Review
U.S. Patent No. 10,297,249

Case Caption	Forum	Patents
		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
<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 Saina Shamilov, Reg. No. 48,266, Brian M. Hoffman, Reg. No. 39,713, 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.

Petition for *Inter Partes* Review
U.S. Patent No. 10,297,249

Dated: July 30, 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 13,629 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: July 30, 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 (1001-1020) are being served via Federal Express on July 30, 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 10,297,249*

Dated: July 30, 2025

FENWICK & WEST LLP

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

Attorneys for Petitioner
Amazon.com Services LLC