

IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC.; AMAZON.COM LLC; AMAZON WEB SERVICES,
INC.; A2Z DEVELOPMENT CENTER, INC. D/B/A LAB126; RAWLES LLC;
AMZN MOBILE LLC; AMZN MOBILE 2 LLC; AMAZON.COM SERVICES,
INC. F/K/A AMAZON FULFILLMENT SERVICES, INC.; AND AMAZON.COM
SERVICES LLC (formerly AMAZON DIGITAL SERVICES LLC),

Petitioners

- vs. -

VB ASSETS, LLC,

Patent Owner

Patent No. 8,073,681

Issued: December 6, 2011

Inventors: Larry Baldwin, Tom Freeman, Michael Tjalve,
Blane Ebersold, Chris Weider

Title: SYSTEM AND METHOD FOR A COOPERATIVE CONVERSATIONAL
VOICE USER INTERFACE

Inter Partes Review No. ____

PETITION FOR *INTER PARTES* REVIEW UNDER 35 U.S.C. §§ 311-319

Mail Stop Patent Board
Patent Trial and Appeal Board
P.O. Box 1450,
Alexandria, VA 22313-1450

July 28, 2020

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. COMPLIANCE WITH REQUIREMENTS FOR A PETITION FOR INTER PARTES REVIEW	1
A. Grounds for Standing (37 C.F.R. § 42.104(a)).....	1
B. Fees for <i>Inter Partes</i> Review (37 C.F.R. § 42.15(a)).....	1
C. Mandatory Notices (37 C.F.R. § 42.8).....	2
III. THRESHOLD FOR REVIEW (35 U.S.C. § 314(a)).....	3
IV. IDENTIFICATION OF CHALLENGED CLAIMS.....	3
A. Relied Upon Patents and Publications.....	3
B. Statutory Grounds.....	4
V. LEVEL OF ORDINARY SKILL IN THE ART.....	4
VI. OVERVIEW OF THE '681 PATENT	5
A. Claim Listing.....	7
VII. CONSTRUCTION OF THE CHALLENGED CLAIMS (37 C.F.R. § 42.104(B)(3))	7
A. “implicit hypothesis” (dependent claims 38, 40, 42)	7
VIII. THE PRIOR ART DISCLOSED THE FEATURES CLAIMED BY THE '681 PATENT	8
A. U.S. Patent App. Pub. No. 2004/0193420 (“Kennewick”)	9
B. Huang, Xuedong et al., “Spoken Language Processing: A Guide to Theory, Algorithm and System Development,” Prentice Hall (2001) (“Huang”)	12

TABLE OF CONTENTS
(Continued)

	Page
C. Seneff, et al., “Hypothesis Selection and Resolution in the Mercury Flight Reservation System” Spoken Language Systems Group, MIT (2001) (“Seneff”).....	12
D. U.S. Patent No. 6,839,896 to Coffman et al. (“Coffman”)	14
E. U.S. Patent App. Pub. No. 2003/0182123 to Mitsuyoshi (“Mitsuyoshi”).....	14
IX. THE CHALLENGED CLAIMS ARE INVALID UNDER § 103.....	15
A. Ground 1: Claims 1-2, 9, 13-14, 21, 25-26, 33, 37, 39, and 41 are rendered obvious by Kennewick.	15
1. Claim 1	15
[1.pre] “A method for providing a cooperative conversational voice user interface, comprising:”	15
[1.a] “receiving an utterance at a voice input device during a current conversation with a user, wherein the utterance includes one or more words that have different meanings in different contexts;”	16
[1.b] “accumulating short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;”	17
[1.c] “accumulating long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;”	18
[1.d] “determining an intended meaning for the utterance,”	19

TABLE OF CONTENTS
(Continued)

	Page
[1.e] “wherein determining the intended meaning for the utterance includes: identifying, at a conversational speech engine, a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge; and”	19
[1.f] “establishing the intended meaning within the identified context, wherein the conversational speech engine establishes the intended meaning within the identified context to disambiguate an intent that the user had in speaking the one or more words that have the different meanings in the different contexts; and”	20
[1.g] “generating a response to the utterance, wherein the conversational speech engine grammatically or syntactically adapts the response based on the intended meaning established within the identified context.”	21
2. Claim 2: “The method of claim 1, wherein accumulating the short-term shared knowledge about the current conversation includes populating a short-term context stack with information about the utterance received during the current conversation.”	23
3. Claim 9: “The method of claim 1, wherein the conversational speech engine grammatically or syntactically adapts the response to influence a subsequent reply utterance that the conversational speech engine expects from the user during the current conversation.”	24
4. Claim 13	25
5. Claim 14	26
6. Claim 21	26

TABLE OF CONTENTS
(Continued)

		Page
7.	Claim 25.....	26
	[25.b] “a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to:”	27
8.	Claim 26.....	28
9.	Claim 33.....	28
10.	Claim 37.....	29
	[37.f] “inferring additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context; and”	29
	[37.g] “establishing the intended meaning within the identified context based on the additional information inferred about the utterance; and”	30
11.	Claim 39.....	31
12.	Claim 41	31
B.	Ground 2: Claims 5, 8, 17, 20, 29, and 32 are rendered obvious by Kennewick in view of Huang.	32
1.	Claim 5: “The method of claim 1, wherein determining the intended meaning for the utterance further includes:”	34
	[5.a] “identifying a conversational goal associated with the utterance, roles associated with the user and one or more other participants in the current conversation, and an information allocation among the user and the one or more other participants in the current conversation; and”	34

TABLE OF CONTENTS
(Continued)

	Page
[5.b] “classifying one or more of the utterance or the current conversation into a conversation type based on one or more of the identified conversational goal, the identified roles, or the identified information allocation, wherein the conversational speech engine further establishes the intended meaning based on the conversation type.”	36
2. Claim 8: “The method of claim 5, wherein the conversational speech engine further grammatically or syntactically adapts the response based on the conversation type.”	38
3. Claim 17	39
4. Claim 20	40
5. Claim 29	40
6. Claim 32	40
C. Ground 3: Claims 10, 22, 34, 38, 40, and 42 are rendered obvious by Kennewick in view of Seneff.	41
1. Claim 10: “The method of claim 1, further comprising:”	43
[10.a] “generating multiple preliminary interpretations of the utterance at a speech recognition engine coupled to the voice input device and the conversational speech engine, wherein an initial interpretation of the utterance comprises one of the multiple preliminary interpretations having a highest confidence level; and”	43

TABLE OF CONTENTS
(Continued)

	Page
[10.b] “updating the short-term shared knowledge about the current conversation to remove the initial interpretation from the multiple preliminary interpretations in response to determining that the initial interpretation was incorrect, wherein the conversational speech engine determines the intended meaning based on one of the multiple preliminary interpretations having a next highest confidence level.”	45
2. Claim 22	47
3. Claim 34	47
4. Claim 38: “The method of claim 37, wherein the established intended meaning comprises an implicit hypothesis having a corresponding degree of certainty about an intent that the user had in speaking the utterance.”	48
5. Claim 40	49
6. Claim 42	49
D. Ground 4: Claims 6-7, 18-19, and 30-31 are rendered obvious by Kennewick and Huang, further in view of Seneff.....	50
1. Claim 6	51

TABLE OF CONTENTS
(Continued)

	Page
2. Claim 7: “The method of claim 6, further comprising generating a preliminary interpretation of the utterance at a speech recognition engine coupled to the voice input device and the conversational speech engine, wherein the conversational speech engine assigns the degree of certainty to the hypothesis based on one or more of the conversation type, information associated with the identified context, or a confidence level associated with the preliminary interpretation generated at the speech recognition engine.”	51
3. Claim 18	52
4. Claim 19	52
5. Claim 30	53
6. Claim 31	53
E. Ground 5: Claims 11-12, 23-24, and 35-36 are rendered obvious by Kennewick in view of Coffman.	53
1. Claim 11: “The method of claim 1, wherein the user speaks the utterance in a multi-modal input that further includes one or more non-voice inputs relating to the utterance.”	55
2. Claim 12: “The method of claim 1, wherein the conversational speech engine generates the response in a multi-modal output that includes one or more non-voice outputs that relate to the utterance or one or more tasks executed to process a request identified from the intended meaning.”	56
3. Claim 23	57
4. Claim 24	58
5. Claim 35	58

TABLE OF CONTENTS
(Continued)

	Page
6. Claim 36.....	58
F. Ground 6: Claims 3-4, 15-16, and 27-28 are rendered obvious by Kennewick in view of Mitsuyoshi.....	58
1. Claim 3: “The method of claim 2, wherein accumulating the short-term shared knowledge about the current conversation further includes expiring the information about the utterance from the short-term context stack after a psychologically appropriate amount of time.”	60
2. Claim 4: “The method of claim 3 wherein accumulating the long-term shared knowledge about the user includes updating one or more long-term profiles associated with the user to include information about the utterance received during the current conversation and relevant data associated with the information expired from the short, term context stack.”	62
3. Claim 15.....	63
4. Claim 16.....	63
5. Claim 27	63
6. Claim 28.....	64
X. THE PETITION SHOULD NOT BE DENIED UNDER §§ 314 OR 325.....	64
XI. CONCLUSION.....	66

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>Amazon.com, Inc., v. M2M Solutions LLC</i> , IPR2019-01204, Paper No. 14 (PTAB Jan. 23, 2020)	65
<i>Amgen Inc. v. Alexion Pharm., Inc.</i> , IPR2019-00739, 2019 WL (PTAB Aug. 30, 2019).....	65
<i>Becton, Dickinson & Co. v. B. Braun Melsungen AG</i> , IPR2017-01586, Paper No. 8 (PTAB Dec. 15, 2017)	64, 65
<i>Ex parte Kubin</i> , 2007 Pat. App. LEXIS 13 (B.P.A.I. May 31, 2007), <i>aff'd</i> , 561 F.3d 1351 (Fed. Cir. 2009).....	52, 57
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (en banc)	7
<i>Realtime Data, LLC v. Iancu</i> , 912 F.3d 1368 (Fed Cir. 2019)	4
STATUTES AND RULES	
35 U.S.C. §§ 102(a) and (b).....	4, 9, 12, 14
35 U.S.C. § 103.....	4, 9, 15
35 U.S.C. § 314.....	3, 64
35 U.S.C. § 325.....	64, 65
Rule 42.104(b)(4)-(5).....	15
OTHER AUTHORITIES	
37 C.F.R. § 42.100(b)	7

EXHIBIT LIST (37 C.F.R. § 42.63(e))

Exhibit	Description
1001	U.S. Patent No. 8,073,681
1002	Declaration of Padhraic Smyth, Ph.D.
1003	U.S. Patent App. Pub. No. 2004/0193420 to Kennewick et al.
1004 – 1007	Reserved
1008	Certified Prosecution History of U.S. Patent No. 8,073,681
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 (Aug. 1994)
1010	Lucente, M., “Conversational interfaces for e-commerce applications,” Communications of the ACM (Sept. 2000), 43(9), 59-61
1011	Huang, et al., “Spoken Language Processing: A Guide to Theory, Algorithm and System Development,” Prentice Hall (2001)
1012	U.S. Patent No. 6,859,776 to Cohen et al.
1013	Curriculum Vitae of Padhraic Smyth, Ph.D.
1014	Declaration of Rachel J. Watters (for Huang, et al., “Spoken Language Processing: A Guide to Theory, Algorithm and System Development,” Prentice Hall (2001))
1015 – 1024	Reserved
1025	Seneff, et al., “Hypothesis Selection and Resolution in the Mercury Flight Reservation System,” Spoken Language Systems Group, MIT (2001)

Exhibit	Description
1026	Declaration of Stephanie Seneff (for Seneff, et al., “Hypothesis Selection and Resolution in the Mercury Flight Reservation System,” Spoken Language Systems Group, MIT (2001))
1027	U.S. Patent No. 6,839,896 to Coffman et al.
1028	U.S. Patent App. Pub. No. 2003/0182123 to Mitsuyoshi
1029	Reserved
1030	Claim Listing for U.S. Patent No. 8,073,681

I. INTRODUCTION

In accordance with 35 U.S.C. §§ 311-319 and 37 C.F.R. §§ 42.1-.80, 42.100-.123, Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., A2Z Development Center, Inc. d/b/a Lab126, Rawles LLC, AMZN Mobile LLC, AMZN Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) (“Petitioners”) respectfully request review of claims 1-42 of U.S. Patent No. 8,073,681 to Baldwin et al., titled “System and Method for a Cooperative Conversational Voice User Interface” (the “’681 patent”), owned by VB Assets, LLC, (“Patent Owner”). Ex. 1001. Because it is reasonably likely that Petitioners will prevail on at least one challenged claim, the Patent Trial and Appeal Board should institute review of the ’681 patent.

II. COMPLIANCE WITH REQUIREMENTS FOR A PETITION FOR *INTER PARTES* REVIEW

A. Grounds for Standing (37 C.F.R. § 42.104(a))

Petitioners certify that the ’681 patent is eligible for *inter partes* review and Petitioners are not barred or estopped from requesting the review.

B. Fees for *Inter Partes* Review (37 C.F.R. § 42.15(a))

The Director is authorized to charge the fee specified by 37 C.F.R. § 42.15(a) to Deposit Account No. 19-2555.

C. Mandatory Notices (37 C.F.R. § 42.8)

Petitioners are the real parties-in-interest. No other party had access to, control over, or contributed to funding of the preparation or filing of this Petition.

Patent Owner asserts the '681 patent against Petitioners¹ in a related case: *VB Assets, LLC v. Amazon.com Inc., et al.*, Civil Action No. 1:19-cv-01410-MN (D. Del. July 29, 2019). Petitioners are unaware of any other judicial or administrative matter that would affect, or be affected by, a decision in this proceeding.

Petitioners designate J. David Hadden, Reg. No. 40,629, as lead counsel, Saina Shamilov, Reg. No. 48,266, and Allen Wang, Reg. No. 68,456, all with Fenwick & West LLP, 801 California Street, Mountain View, CA 94041, telephone: (650) 988-8500, fax: (650) 938-5200.

Service of any documents may be made to the postal mailing address above. Petitioners consent to electronic service at VBAssets-IPR@fenwick.com. A Power of Attorney is being filed concurrently with the designation of counsel in accordance with 37 C.F.R. § 42.10(b).

¹ Amazon Digital Services LLC, named in the D. Del. case, became Amazon.com Services LLC on January 1, 2020. Neither Rawles LLC nor AMZN Mobile 2 LLC are properly named defendants.

III. THRESHOLD FOR REVIEW (35 U.S.C. § 314(a))

It is reasonably likely that Petitioners will prevail on at least one of the challenged claims because the subject matter recited in claims 1-42 of the '681 patent was well known in the prior art, as shown herein. Any motivation to combine the prior art is provided herein as necessary.

IV. IDENTIFICATION OF CHALLENGED CLAIMS

A. Relied Upon Patents and Publications

- Ex. 1003: U.S. Patent App. Pub. No. 2004/0193420 (“Kennewick”)
- Ex. 1011: Huang, Xuedong et al., “Spoken Language Processing: A Guide to Theory, Algorithm and System Development,” Prentice Hall (2001) (“Huang”)²
- Ex. 1025: Seneff, Stephanie et al., “Hypothesis Selection and Resolution in the Mercury Flight Reservation System” Spoken Language Systems Group, MIT (2001) (“Seneff”)³
- Ex. 1027: U.S. Patent No. 6,839,896 to Coffman et al. (“Coffman”)
- Ex. 1028: U.S. Patent App. Pub. No. 2003/0182123 to Mitsuyoshi (“Mitsuyoshi”)

² Huang was publicly available by November 20, 2002. Ex. 1014.

³ Seneff was publicly available by March 21, 2001. Ex. 1026.

B. Statutory Grounds

Petitioners request *inter partes* review on the grounds below under 35 U.S.C. § 103.

Ground	35 U.S.C.	References	Claims Challenged
1	§ 103 ⁴	Kennewick	1-2, 9, 13-14, 21, 25-26, 33, 37, 39, and 41
2	§ 103	Kennewick and Huang	5, 8, 17, 20, 29, and 32
3	§ 103	Kennewick and Seneff	10, 22, 34, 38, 40, and 42
4	§ 103	Kennewick, Huang, and Seneff	6-7, 18-19, and 30-31
5	§ 103	Kennewick and Coffman	11-12, 23-24, and 35-36
6	§ 103	Kennewick and Mitsuyoshi	3-4, 15-16, and 27-28

V. LEVEL OF ORDINARY SKILL IN THE ART

A “person of ordinary skill in the art” (“POSITA”) would have at least a Bachelor-level degree in computer science, computer engineering, electrical

⁴ Ground 1 includes anticipation based on Kennewick: “it is well settled that a disclosure that anticipates under § 102 also renders the claim invalid under §103, for anticipation is the epitome of obviousness.” *Realtime Data, LLC v. Iancu*, 912 F.3d 1368, 1373 (Fed Cir. 2019) (internal citations and quotations omitted) (no error in § 103 invalidity based on a single reference).

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. Ex. 1002, ¶¶28-31.

VI. OVERVIEW OF THE '681 PATENT

The '681 patent was filed October 16, 2006 and does not claim priority to any earlier application.⁵

The '681 patent, titled “System and Method for a Cooperative Conversational Voice User Interface,” generally describes a “human to machine voice user interface” that uses short-term and long-term knowledge to understand and respond to utterances in an identified context. *See, e.g.*, Ex. 1001, 3:15-18; 3:41-47; 4:33-38; 4:45-53. “Short-term knowledge may accumulate during a single conversation,” *e.g.*, as a “history of the conversation.” *Id.*, 4:53-55, 16:3-7. “Long-term knowledge” is “unconstrained by a current session,” is “user-centric,” and “accumulate[s] over time to build user profiles.” *Id.*, 5:4-8, 13:34-37, 16:29-34. The claims of the '681 patent are directed to methods and systems for providing a “cooperative conversational voice user interface,” using voice input devices to

⁵ Patent Owner has not claimed an earlier priority date in the district court proceeding.

receive utterances, determine intended meaning based on these knowledge sources, and generate a response within the identified context. *Id.*, 19:35-65.

The '681 patent describes receiving as input a human utterance, which may be accompanied by other non-voice inputs. *Id.*, 2:6-14. A speech recognition engine generates preliminary interpretations of the utterance, which a conversational speech engine processes further to generate an “adaptive conversational response.” *Id.*, 2:16-25. To identify “all relevant information for completing one or more tasks,” the adaptive response prompts the user by asking for missing context information “in a way that most restricts possible interpretations.” *Id.*, 2:49-64.

The '681 patent further describes determining context using short-term knowledge from a current conversation and long-term knowledge that is “user-centric, rather than session-based.” *Id.*, 4:51-55; 5:4-8; 14:3-4. “Context domain agents” “disambiguate” one context from another and assign “degrees of certainty” to generate hypotheses about a user’s intent. *Id.*, 12:19-59. The '681 patent further describes using the type of conversation to generate hypotheses, based on conversational goals, participant roles, or information allocation among participants. *Id.*, 14:31-39; 15:55-59. Short-term and long-term knowledge assist in creating hypotheses of a user’s intent when necessary information is missing. *Id.*, 15:65-16:53. The '681 patent further describes using context, task recognition information,

and knowledge about a user to create “conversational feel” in system responses. *Id.*, 17:31-18:44.

A. Claim Listing

Exhibit 1030 is a claim listing enumerating each claim element.

VII. CONSTRUCTION OF THE CHALLENGED CLAIMS (37 C.F.R. § 42.104(B)(3))

Claim constructions must comport with “intrinsic” evidence: the words of the claims, the specification, the figures, and the patent prosecution history. *See* 37 C.F.R. § 42.100(b) (same standard for IPR and civil action); *Phillips v. AWH Corp.*, 415 F.3d 1303, 1314-17 (Fed. Cir. 2005) (en banc). Terms should be given the meaning “a person of ordinary skill in the art in question at the time of the invention” would ascribe them based on the patent and its file history. *Phillips*, 415 F.3d at 1313. For this petition, the Board should adopt the construction described below.⁶

A. “implicit hypothesis” (dependent claims 38, 40, 42)

In the ’681 patent, a hypothesis refers to a potential “intent of a user utterance.” Ex. 1001, Abstract, 5:48-50. The system uses these hypotheses of user intent to generate appropriate responses. *Id.*, 5:61-66.

⁶ For the avoidance of any doubt, nothing in this Petition is a waiver of any arguments, positions, or appellate rights of Petitioners, including, but not limited to, claim construction.

The '681 patent differentiates “explicit” and “implicit” hypotheses. A hypothesis builder “may generate one or more explicit hypotheses of a user’s intent when an utterance contains all information (including qualifiers) needed to complete a request or task,” while it may “generate one or more implicit hypotheses of a user’s intent when an utterance may be missing required qualifiers or other information needed to complete a request or task.” *Id.*, 15:55-59, 15:65-16:2. To generate the implicit hypothesis, the hypothesis builder may “supplement or infer additional information” based on short-term or long-term contextual knowledge. *Id.*, 16:3-10, 16:24-28. Thus, a POSITA would have understood “implicit hypothesis” to mean a potential intent of a user utterance generated by filling in missing information using contextual data. Ex. 1002, ¶¶25-27.

VIII. THE PRIOR ART DISCLOSED THE FEATURES CLAIMED BY THE '681 PATENT

The concepts claimed in the '681 patent for a natural language speech interface were well known to POSITAs and were described in patents and publications before the October 2006 priority date. The relevant technological background known to POSITAs at the time included automated speech recognition, natural language understanding, dialog management, and user modeling (by these and other names). Ex. 1002, ¶¶35-48. The references below discuss interactive speech systems and disclose methods for using context, prior information and user data to provide accurate responses. A POSITA would have known that

improvements to speech interfaces in one application could be adapted for use in other applications in a way that would achieve similar beneficial results. *Id.*, ¶48.

POSITAs were aware that natural language processing and speech recognition could be improved by using domain knowledge, histories, and user profiles to prepare and evaluate hypotheses regarding user intent, and responding accurately, in a natural language style, to user queries. *See, e.g., id.*, ¶49; Ex. 1003, [0005], [0011]; Ex. 1011, 898, 854; Ex. 1025, Abstract; Ex. 1027, 9:26-28, 9:46-49, 21:44-22:19.

A. U.S. Patent App. Pub. No. 2004/0193420 (“Kennewick”)

Kennewick, published September 30, 2004, is prior art to the ’681 patent under 35 U.S.C. §§ 102(a) and (b). Both Kennewick and the patent application that issued as the ’681 patent, which does not claim priority to Kennewick, were filed by VoiceBox Technologies Corporation (“VoiceBox”),⁷ but this common ownership does not disqualify Kennewick because it is § 102(b) art. *See* § 103(c)(1) (commonly owned “prior art *only under* one or more of subsections (e), (f), and (g) of section 102, shall not preclude patentability under this section”).⁸

Kennewick describes a system that receives natural language input and “present[s] results in a natural manner.” Ex. 1003, [0002]. The system is “a

⁷ VoiceBox is a predecessor-in-interest of Patent Owner.

⁸ Unless otherwise noted, all emphasis is added.

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.*, [0011]. The system includes “a speech interface device that receives spoken natural language queries, commands and/or other utterances from a user, and a computer device or system that receives input from the speech unit and processes the input (*e.g.*, retrieves information responsive to the query, takes action consistent with the command and performs other functions as detailed herein), and responds to the user with a natural language speech response.” *Id.*, [0012]-[0014].

Kennewick explains that a “dialog with the system may begin when a user first addresses it” and may end when the user “speak[s] a dismissal word (‘good by’ [*sic*]).” *Id.*, [0156]. The speech recognition engine may determine the user’s identity “using unique characteristics of their speech” and tag “[r]ecognized words and phrases” with this identity in all further processing allowing “other components of the system to tie that utterance to the correct user and dialog.” *Id.*, [0155]-[0156]. Since “[m]ultiple users can use the system,” recognizing the user allows the system to “determine which user is stating each question or command and apply the correct profile and context.” *Id.*, [0037], [0119].

To process utterances, Kennewick describes “determin[ing] a context for an utterance by applying prior probabilities or fuzzy possibilities to” a “dialog history,” among other things. *Id.*, [0160], [0164]. As Kennewick explains, “some questions can only be interpreted in the context of previous questions.” *Id.*, [0007]. For example, in one dialog, “a question with the keywords ‘temperature’ implies a context value of weather for the question,” while in “a different dialog, the keyword ‘temperature’ can imply a context for a measurement.” *Id.*, [0160]. To facilitate context determination, Kennewick also teaches “creat[ing], stor[ing] and us[ing] extensive personal profile information for each user,” “improving the reliability of determining the context and presenting the expected results for a particular question or command.” *Id.*, Abstract, [0031], [0160] (describing “determin[ing] a context for an utterance” based on a “user profile,” among other values). Kennewick discloses that its “user profile” includes, among other things, “history of questions asked,” “session histories,” “history of commands and queries,” and “history of [the user’s] interactions with the system.” *Id.*, [0031], [0143]. Of the possible contexts determined based on the dialog history and user profile, Kennewick scores and uses the “top one or few” for further processing to determine the meaning of each utterance. *Id.*, [0160]-[0162].

B. Huang, Xuedong et al., “Spoken Language Processing: A Guide to Theory, Algorithm and System Development,” Prentice Hall (2001) (“Huang”)

“Spoken Language Processing: A Guide to Theory, Algorithm and System Development” by Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon (“Huang”), was published in 2001 by Prentice Hall and is prior art to the ’681 patent under 35 U.S.C. §§ 102(a) and (b). Huang discloses techniques for improving voice recognition technology to better recognize and interpret natural language voice-based inputs. Ex. 1011, Preface, Introduction. The speech processing systems taught in Huang include components for understanding spoken natural language utterances and managing dialog. *Id.*, 853-854, 886.

Huang discloses that spoken language meaning is connected to the goals of a listener and speaker (“participant”) in the interaction. *Id.*, 854, 856. A model of conversational “participants’ goals, commitments, and beliefs” can be classified into a “plan” using which the system can improve its cooperative nature and utterance interpretation abilities. *Id.*, 890-891.

C. Seneff, et al., “Hypothesis Selection and Resolution in the Mercury Flight Reservation System” Spoken Language Systems Group, MIT (2001) (“Seneff”)

Seneff, published 2001, is prior art to the ’681 patent under 35 U.S.C. §§ 102(a) and (b).

Seneff discloses a spoken dialogue system, the MIT Mercury flight reservation system, in which a dialogue manager uses local information, global knowledge, and specific knowledge about a user to select the best interpretation of a human-to-machine utterance. Ex. 1025, Abstract, §1. Seneff describes a “more natural interface” that achieves better understanding of user queries through a hypothesis selection process. *Id.*, §1. Seneff’s Mercury system uses a dialog history of the current conversation for hypothesis selection. *Id.*, §§1, 2.2-2.3. Seneff’s system selects a preferred hypothesis of the user’s intent, using “confidence scores,” consistency with the “prior dialogue,” and other knowledge sources, in a process illustrated in Figure 4 below. *Id.*, §§2.2, 4. The knowledge sources track the state of the dialog as it progresses, including “any preceding system-initiated query as well as all user-specified constraints.” *Id.*, §4.

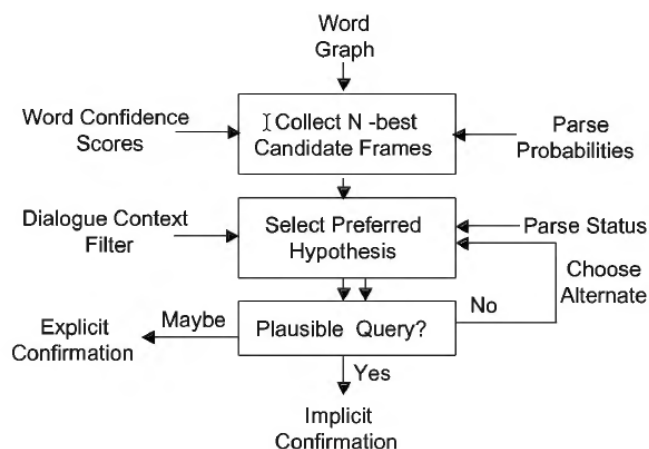


Figure 4: A block diagram of the process of hypothesis selection and verification.

D. U.S. Patent No. 6,839,896 to Coffman et al. (“Coffman”)

Coffman, issued on January 4, 2005, is prior art to the ’681 patent under 35 U.S.C. §§ 102(a) and (b).

Coffman discloses a system and method for understanding and responding to natural language utterances that may be accompanied by other non-voice inputs. Ex. 1027, 4:17-30. To do so, Coffman teaches using “short term and long term context (including previous input and output) to disambiguate and understand the user’s intention.” *Id.*, 4:9-16. Coffman’s context resolution system for disambiguating utterances ultimately “generates an interpretation of the user’s intent” to which it responds. *Id.*, 8:10-16

E. U.S. Patent App. Pub. No. 2003/0182123 to Mitsuyoshi (“Mitsuyoshi”)

Mitsuyoshi, published September 25, 2003, is prior art to the ’681 patent under 35 U.S.C. §§ 102(a) and (b).

Mitsuyoshi discloses a voice recognition and natural language processing system and method. Ex. 1028, [0066]. Mitsuyoshi teaches storing historical knowledge in a “long-term memory” (long-term knowledge) and “newly inputted” information about a current conversation or “episode” (short-term knowledge). *Id.*, [0060], [0168], [0215]. Over time, the system may cycle out unrequired information that the system deems less important. *Id.*, [0228] (prioritizing retention of “knowledge used repeatedly” and “data determined to have a strong impression”).

Mitsuyoshi teaches that the disclosed methods of knowledge management can be used to configure a “variety of systems, each including a dialog function between a computer and a human.” *Id.*, [0264].

IX. THE CHALLENGED CLAIMS ARE INVALID UNDER § 103.

Pursuant to Rule 42.104(b)(4)-(5), the following analysis demonstrates where every element of the challenged claims is disclosed by each ground.

A. Ground 1: Claims 1-2, 9, 13-14, 21, 25-26, 33, 37, 39, and 41 are rendered obvious by Kennewick.

Kennewick, alone or in view of the knowledge of a POSITA, renders obvious claims 1-2, 9, 13-14, 21, 25-26, 33, 37, 39, and 41.⁹ Ex. 1002, ¶67.

1. Claim 1

[1.pre] “A method for providing a cooperative conversational voice user interface, comprising:”

To the extent the preamble is limiting, Kennewick discloses a “complete speech-based natural language query and response environment” that “processes input” and “responds to the user with a natural language speech response.” Ex. 1003, Abstract, [0014]. When responding, the speech interface “format[s] results in a manner enhancing the understandability to the user” based on context, interaction history, and domain. *Id.*, [0023]. The speech interface avoids formulating responses

⁹ See *supra* note 4.

in a “rigid” or “unnatural” way and offers the user control over how information is presented. *Id.*, [0024]-[0025]. Thus, Kennewick discloses the claimed “cooperative conversation voice user interface.” *Id.*, [0023]; Ex. 1002, ¶68.

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

Kennewick discloses receiving an utterance during a current conversation (or “dialog”) that begins “when a user first addresses it” and ends with a “dismissal word.” Ex. 1003, [0156]. The received utterances can contain words that mean different things based on the context of the utterance in a conversation. For example, Kennewick teaches that in one dialog “a question with the keywords [*sic*] ‘temperature’ implies a context value of *weather* for the question. Within a different dialog, the keyword “temperature” can imply a context for a *measurement*.” *Id.*, [0160]. That is, “temperature” could be interpreted as an outdoor temperature or as a body temperature depending on the context. Ex. 1002, ¶69. The same can be true for multiple words: in another example, Kennewick uses “flight information in the database and network information along with context” to resolve the ambiguous query “what about flight one hundred and twenty too.” Ex. 1003, [0163] (possible meanings being “flight 100 and flight 20 also, flight 100 and flight 22, flight 122, etc.”). A context in which “flight 122” exists would result in a different meaning than a context where the system just listed several flight options including “flight

100” and “flight 20.” Ex. 1002, ¶70. Kennewick thus discloses receiving utterances that “include[] one or more words that have different meanings in different contexts.”

[1.b] “accumulating short-term shared knowledge about the current conversation, wherein the short-term shared knowledge includes knowledge about the utterance received during the current conversation;”

Kennewick discloses accumulating short-term knowledge about the current dialog with a user. The system first “determine[s] the user’s identity by voice and name for each utterance” and then adds corresponding “tags ... to each utterance.” Ex. 1003, [0155]. Because the system tags “[r]ecognized words and phrases ... with this identity in all further processing,” other system components can “tie that utterance,” and thus the corresponding words and phrases, “to the correct user and dialog.” *Id.* This allows the system to fill in information required to clarify and fully specify the user’s intent where the user’s utterance does not unambiguously include all required information. *Id.*, [0164]. In Kennewick, the “history of *the* dialog” reflects the current conversation; a “dialog” begins when a user first addresses the system and ends with a dismissal word. *Id.*, [0156], [0170].

Kennewick discloses “fuzzy set probabilities or prior possibilities” may be derived from “the history of the dialog,” among other things. *Id.*, [0162] (“based on user responses,” updating the probabilities/possibilities “as the system learns the desired behavior”), [0157] (“dynamically updat[ing]” the the

probabilities/possibilities based on “user dialog history”). Dynamically updating of the probabilities and possibilities “maximize[s] the probability of correct recognition *at each stage of the dialog.*” *Id.*

Thus, a POSITA would understand that Kennewick’s system of tagging recognized words and phrases of each utterance with a user identity allows the system to build a dialog history of words and phrases uttered by that user during the conversation. Ex. 1002, ¶¶71-73. Kennewick’s description of building dialog history by tagging dialog utterances is consistent with the ’681 patent’s disclosure of “short-term knowledge” including a “history of the conversation” built “as the conversation progresses.” Ex. 1001, 15:65-16:7, 15:38-42. The dialog history therefore accumulates “short-term shared knowledge about the current conversation.”

[1.c] “accumulating long-term shared knowledge about the user, wherein the long-term shared knowledge includes knowledge about one or more past conversations with the user;”

Kennewick “creates, stores and uses extensive personal profile information for each user,” including the “history of questions asked,” “session *histories*,” “history of commands and queries,” and “history of [individual] interactions with the system.” Ex. 1003, [0031], [0143]. A POSITA would understand that the user profile thus accumulates long-term knowledge of the user’s history of interaction

with the system unconstrained by the current conversation. Ex. 1002, ¶74. Therefore, Kennewick discloses this limitation.

This understanding of “long-term shared knowledge” is consistent with how the ’681 patent describes it. The ’681 patent also discloses “long-term knowledge may accumulate over time to build user profiles” that are “unconstrained by a current session.” Ex. 1001, 13:34-37, 16:29-32; Ex. 1002, ¶75.

[1.d] “determining an intended meaning for the utterance,”

Kennewick discloses a multi-step process to interpret each utterance and determine the intended meaning by resolving “inconsistent, ambiguous, conflicting and incomplete information.” Ex. 1003, [0028]-[0029], [0162]-[0165]. These steps, described further regarding limitations [1.e]-[1.f] below, effectively map the informal natural language utterance of the user to a well-formed question or command that the system can process. *Id.*, [0165]-[0169]; Ex. 1002, ¶76.

[1.e] “wherein determining the intended meaning for the utterance includes: identifying, at a conversational speech engine, a context associated with the utterance from the short-term shared knowledge and the long-term shared knowledge; and”

“The context and expected results from a particular question may be highly dependent on the individual asking the question.” Ex. 1003, [0031]. Identifying a context may help resolve ambiguity; as described in Section IX.A.1[1.a], the word “temperature” may be understood differently in the contexts of “weather” or “measurement.” Identifying the context—weather, measurement, etc.—permits

assigning a meaning to the utterance. Ex. 1002, ¶77; Ex. 1003, [0078]-[0081], [0160]. As another example, the system uses context to determine the intended meaning of the utterance “what about flight one hundred and twenty too?” among the possibilities of “flight 100 and flight 20 also, flight 100 and flight 22, flight 122, etc.” Ex. 1003, [0163].

Thus, to determine the intended meaning, a parser coupled to the speech recognition engine “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.” *Id.*, [0160]. As described in Sections IX.A.1[1.b] and [1.c], the “dialog history” includes short-term shared knowledge and the “user profile” includes long-term shared knowledge. The parser dynamically receives these probabilities or possibilities to score the possible contexts for further processing. *Id.*, [0160]-[0161]. Thus, Kennewick discloses this limitation. Ex. 1002, ¶78.

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

Again, as described regarding limitation 1.e above, the Kennewick system uses context to determine an intended meaning, which can disambiguate the user’s intent in speaking the words of the utterance. In the temperature example, the system

may use a weather context to understand that the user is requesting a forecast and not their body temperature. Ex. 1003, [0160]; Ex. 1002, ¶79. In the flight example, the system may use the context of the dialog and information about actual flight numbers to interpret the ambiguous query “what about flight one hundred and twenty too?” Ex. 1003, [0163]. For instance, the flight reservation context may result in a greater likelihood of interpreting the recognized word “too” as the number “two.” Ex. 1002, ¶79. These illustrate the system disambiguating the user’s intent based on “context of the domain of expertise required,” “the context for the question or command,” “domain specific information,” “profile data,” and “user preferences,” among other information. Ex. 1003, [0029], [0031]-[0032]; *see also* Section IX.A.1[1.a] (regarding “words that have different meanings in different contexts”).

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

Kennewick’s system, through the “significant use of context,” creates a “natural environment” to “present[] expected results for a particular question or command” in the format of natural language speech responses. Ex. 1003, Abstract, [0014]. As described in relation to the preceding claim limitations, the system uses an identified context to discern the intended meaning of an utterance.

The system then prepares a response using an agent corresponding to the context, *id.*, [0162], which queries information sources and scores their relevance to

determine “if a single best answer can be extracted.” *Id.*, [0168], [0170]-[0171]. The agent evaluates results until a satisfactory response is constructed based on context-specific tokens: “[f]or example, for a question on weather, the tokens can include the date, day of week, predicted high temperature, predicted low temperature, chance of precipitation, expected cloud cover, expected type of precipitation and other tokens.” *Id.*, [0171], [0175]. Then, the agent formats the presentation of a response “based on the context, the criteria, and the user profile” to provide a response that is “best understood and most natural to the user.” *Id.*, [0178]. In formatting the response, the agent “may perform variable substitutions and transformations,” “vary the order of presentation ... and the exact terminology used,” and “select the presentation personality.” *Id.* Because the resulting response is formatted based on the user’s request to ask about the weather, Kennewick thus discloses adapting the response based on the intended meaning, which can “enhance[e] the understandability to the user.” *Id.*, [0023].

Based on the disclosures above, a POSITA would understand that Kennewick’s method of formatting responses, including through terminology selection and choice of presentation personality, thus discloses “grammatically or syntactically adapt[ing]” responses based on the meaning of the utterance within the context as recited by claim 1. Ex. 1002, ¶¶80-82. The disclosure is consistent with the description in the ’681 patent of using “syntactically, grammatically, and

contextually sensitive ‘intelligent responses’” to “generate a conversational experience for a user, while also guiding the user to reply in a manner favorable for recognition.” Ex. 1001, 17:21-26.

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

Kennewick discloses “accumulating short-term shared knowledge” of the current conversation as a dialog history. See Sections IX.A.1[1.a], [1.b]. As taught by Kennewick, “[p]art of the dialog history is maintained in a context stack.” Ex. 1003, [0170]. Thus, at least part of the stack is a “short-term context stack” populated with information about utterances from the current conversation. Maintaining this information permits, for example, inheritance of “key words and criteria” into abbreviated user questions when the “criteria of the question remain constant.” *Id.*, [0153]. Using a context stack was a known, standard way of storing short-term information from a current conversation at the time. Ex. 1002, ¶83. A POSITA would have known it was also common to track long-term knowledge across conversations in another way. *Id.* Thus, Kennewick discloses this limitation.

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

Kennewick discloses asking verifying questions “phrased to indicate the context of the question including all criteria or parameters.” Ex. 1003, [0033]. “For example, the question can be in the form of: ‘Did I understand that you want such-and-such.’ If the user confirms that the question is correct, the system may proceed to produce a response.” *Id.*, [0161]. A POSITA would recognize “such-and-such” in this example as a placeholder to be replaced with a grammatically and syntactically appropriate version of the recognized question, command, or topic to influence a subsequent confirming reply from the user. Ex. 1002, ¶84. For example, for an utterance recognized as “Get me a flight to Seattle,” the system may fill in criteria and rephrase the request as “Did I understand that you want me to look for flights from Dallas to Seattle?” *Id.* Alternatively, if the system cannot infer an answer based on the information provided, it can ask for more information through the speech interface. Ex. 1003, [0171]-[0172]. In each instance, the system adapts these verification questions with a goal of obtaining a clarifying reply from the user. Ex. 1002, ¶84.

4. Claim 13

Claim 13 is substantially similar to claim 1, but adds “non-transitory computer readable medium containing computer-executable instructions” and “computer-executable instructions operable when executed” in preamble 13.pre and minor variations in syntax/tense in limitations 13.a-13.f. The ’681 patent does not describe or distinguish a particular computer readable medium or computer-executable instructions; thus 13.pre is obvious for the reasons in Section IX.A.1[1.pre].

To the extent 13.pre is limiting, Kennewick discloses both handheld and fixed computer embodiments with “processing units” that use “nonvolatile memory,” such as “flash memory and hard disk drives.” Ex. 1003, [0111], [0115], Figs. 3, 4. The disclosed computers use these non-transitory memories to store “software and data.” *Id.*, [0111], [0115]. A POSITA would understand that the stored software includes the modules and agents that perform the functions of the Kennewick system. *See, e.g., id.*, [0013]-[0014] (identifying, for example, a “speech recognition module,” “parser,” “network interface,” “graphical user interface,” and more, all of which “can be distributed in any way between a handheld device, a computer attached to a vehicle, a desktop computer or a server without altering the function, features, scope, or intent of the invention”), [0017] (“[a]gents are executables”); Ex. 1002, ¶¶85-86.

Kennewick thus discloses a non-transitory computer readable medium containing computer-executable instructions, *i.e.*, software. A POSITA would

recognize that these instructions, when executed by the computer, are operable to perform the various functions of the Kennewick system, including those cited in relation to the limitations of claim 1. Ex. 1002, ¶¶87-93. Thus, Kennewick renders claim 13 obvious for the reasons in Section IX.A.1, with limitations 13.a-13.c and 13.d-13.f corresponding to limitations 1.a-1.c and 1.e-1.g, respectively.

5. Claim 14

The additional limitation of this dependent claim is substantially similar to that of claim 2, other than reciting “the computer-executable instructions are further operable when executed to” and is obvious for the reasons in Sections IX.A.2 (claim 2) and IX.A.4[13.pre] (computer-executable instructions). Ex. 1002, ¶94.

6. Claim 21

The additional limitation of this dependent claim is substantially similar to that of claim 9, other than reciting “adapts more the response” instead of “adapts the response.” As a POSITA would have understood that adaptation of a response as disclosed by Kennewick could be performed to a lesser or greater degree, the same disclosures apply here as well to the extent this limitation requires “more” of an adaptation. Ex. 1002, ¶95. Thus, this limitation is obvious for the reasons in Section IX.A.3.

7. Claim 25

Claim 25 is a system variant of claim 1, reciting a “system for providing a cooperative conversational voice user interface.” As described in Section

IX.A.1[1.pre], Kennewick discloses a method of providing a “cooperative conversational voice user interface”; the interface is provided by a “system.” Ex. 1003, [0012]-[0014]. Thus, Kennewick discloses preamble 25.pre to the extent it is limiting.

Limitation 25.b recites a “conversational speech engine ... configured to” perform the functions of the remaining limitations of claim 25, paralleling those of claim 1. These remaining limitations are substantially similar to those of claim 1 other than minor syntax/tense variations and factoring out recitations to the “conversational speech engine.” Thus, Kennewick renders claim 25 obvious for the reasons below regarding 25.b and those in Section IX.A.1, with limitations 25.a and 25.c-25.g corresponding to limitations 1.a and 1.b, 1.c, 1.e-1.g, respectively. Ex. 1002, ¶¶96-104.

[25.b] “a conversational speech engine, wherein the conversational speech engine includes one or more processors configured to:”

Kennewick discloses a “main speech processing unit 98” and “speech unit 128” which collectively perform the functions recited in limitations 25.c-25.g. Ex. 1003, [0111], [0115]; *see also id.*, [101] (identifying elements 98 and 128 as “the invention”). These two units may be embedded into handheld or fixed computers, which include “processing units” that in turn “include one or more central processing units,” *i.e.*, processors. *Id.*, [0111], [0115]. Both the main speech processing unit and speech unit may be integrated with these processing units of the computers. *Id.*,

[0122] (“[t]he speech unit can be integrated with the main [speech processing] unit”), [0111], [0115] (“the main speech processing unit 98 can be integrated with the one or more processing units”). Indeed, as Kennewick explains, the “exact distribution and packaging of the main speech processing unit 98 and speech unit 128 ... will not change the functionality, capabilities or spirit of the invention in any way.” *Id.*, [0109]. Thus, a POSITA would have understood that Kennewick discloses a “conversational speech engine” including “one or more processors” configured to perform the limitations of claim 25. Ex. 1002, ¶¶98-99.

8. Claim 26

Claim 26 is similar to claim 14, other than replacing “computer-executed instructions are further operable when executed to” with “one or more processors are further configured to” and is obvious for the reasons in Section IX.A.5. A POSITA would have understood that the processors described in Section IX.A.7[25.b], which perform the function of “accumulating short-term shared knowledge,” Section IX.A.7[25.c], do so by executing the computer-executed instructions described regarding claim 14. Ex. 1002, ¶105.

9. Claim 33

Claim 33 is similar to claim 9 other than replacing “the conversational speech engine ... adapts the response” with “the one or more processors are further configured to generate the ... adapted response” and omitting the “conversational

speech engine” and is obvious for the reasons in Section IX.A.3. A POSITA would have understood that the processors described in Section IX.A.7[25.b], which perform the function of “generat[ing] a grammatically or syntactically adapted response,” Section IX.A.7[25.g], would also be configured to perform the rest of this limitation for the reasons described regarding claim 9. Ex. 1002, ¶106.

10. Claim 37

Other than limitations 37.f and 37.g, described below, the remaining limitations of claim 37 are either identical to those of claim 1 (37.pre and 37.b-37.e corresponding to 1.pre and 1.b-1.e, respectively) or substantially similar to but broader: 37.a lacks the limiting language “wherein the utterance includes ...” of 1.a and 37.h lacks the limiting language “wherein the conversational speech engine grammatically or syntactically adapts the response” of 1.g. Thus, Kennewick renders claim 37 obvious for the reasons below regarding 37.f and 37.g and those in Section IX.A.1 for the corresponding limitations. Ex. 1002, ¶¶107-116.

[37.f] “inferring additional information about the utterance from the short-term shared knowledge and the long-term shared knowledge in response to determining that the utterance contains insufficient information to complete a request in the identified context; and”

Kennewick discloses that additional criteria may need to be inferred because a user may not “explicitly suppl[y]” all criteria necessary to eliminate ambiguity in a command. Ex. 1003, [0162]-[0163]. Kennewick further discloses inferring “missing values” based on “command context, the user profile 110, command

history,” and “other factors.” *Id.*, [0177]. If available information does not adequately answer a user’s question, Kennewick’s system can ask more questions before formulating an adequate response. *Id.*, [0032]. Even when making additional queries, the system still uses context and user profile information to make other inferences “to minimize the interaction with the user.” *Id.*, [0032], Abstract (system “creates, stores and uses extensive personal profile information for each user, thereby improving the reliability of determining the context and presenting the expected results for a particular question or command”).

A POSITA would thus understand that Kennewick infers additional information, such as criteria and missing values, using short-term knowledge in the context (dialog history, Section IX.A.1[1.b]) and long-term knowledge in the context (user profile, Section IX.A.1[1.c]) to resolve information deficits in each utterance. Ex. 1002, ¶¶113-114.

[37.g] “establishing the intended meaning within the identified context based on the additional information inferred about the utterance; and”

To infer the additional information as described regarding the previous limitation, the parser may use criteria handlers after it identifies the context for the utterance. Ex. 1003, [0162]. These handlers provide “context sensitive procedures for extracting the criteria or parameters” from “sources including, for example, the history of the dialog, the user profile 110, and the agent.” *Id.* The extracted values serve to establish the intended meaning in the identified context. In the weather

context, “examples of criteria include, location, date and time.” *Id.*, [0162]. For instance, the user might ask, “when will it start raining today?” A POSITA would understand that the system would have to infer additional information about location, date, and time from the context to establish the user’s intended meaning before it could respond. Ex. 1002, ¶115. Thus, Kennewick discloses this limitation.

11. Claim 39

Claim 39 recites limitations substantially similar to those of claims 1, 13, and 37; thus, Kennewick renders claim 39 obvious for the reasons described in Sections IX.A.1, IX.A.4, and IX.A.9 for the following limitations: preamble 39.pre is identical to 13.pre; 39.a is substantially similar to 1.a but broader as it lacks the limiting language “wherein the utterance includes ...”; 39.b-39.c are substantially similar to 1.b-1.c other than minor syntax/tense variations; 39.d is substantially similar to 1.e but rephrases “identifying, at a conversational speech engine, a context associated with the utterance” as “identify a context associated with the utterance, wherein a conversational speech engine identifies the context associated with the utterance”; and 39.e-39.g are substantially similar to 37.f-37.h other than minor syntax/tense variations. Ex. 1002, ¶¶117-124.

12. Claim 41

Claim 41 recites limitations substantially similar to those of claims 1, 25, and 37; thus, Kennewick renders claim 41 obvious for the reasons described in Sections

IX.A.1, IX.A.7, and IX.A.9 for the following limitations: preamble 41.pre is identical to 25.pre; 41.a is substantially similar to 1.a but broader as it lacks the requirement that the utterance include words with “different meanings in different contexts”; 41.b is identical to 25.b; 41.c-41.e are substantially similar to 1.b-1.c and 1.e other than minor syntax/tense variations; and 41.f-41.h are substantially similar to 37.f-37.h other than minor syntax/tense variations. Ex. 1002, ¶¶125-133.

B. Ground 2: Claims 5, 8, 17, 20, 29, and 32 are rendered obvious by Kennewick in view of Huang.

As described in Section IX.A, Kennewick renders obvious independent claims 1, 13, and 25, from which claims 5, 8, 17, 20, 29, and 32 depend. This section identifies additional disclosures of Huang that render these dependent claims obvious. Ex. 1002, ¶134.

As described in Section VIII, a POSITA would have considered prior art references in the shared fields of speech-based natural language query and response environments, and spoken language understanding and processing, and could have combined their teachings in a modular way.

Kennewick’s natural language speech processing system uses dialog history (among other sources) as contextual information for understanding an utterance. Ex. 1003, [0160] (parser may determine context using dialog history and context stack). For example, Kennewick discloses using the dialog history to substitute words with corresponding values that the agents can process. *Id.*, [0164]. Kennewick also

dynamically updates words to “maximize the probability of correct recognition at each stage of the dialog,” based on the user dialog history. *Id.*, [0157].

Huang similarly describes using “dialog history” as part of a “discourse analysis” process to improve interpretation accuracy. Ex. 1011, 881. The discourse analysis process involves deciphering relative expressions (*e.g.*, translating the word “tomorrow” into 1/5/2000), inferring missing information, and detecting ambiguities or inconsistencies. *Id.*, 881. In Huang, the dialog history factors into determining the goal of a conversation. *Id.*, 890. Huang discloses a “plan-based” system that uses the determined goal and a model of conversational participants’ beliefs to formulate a corresponding “plan.” *Id.*, 888-889. The plan provides a beneficial contextual tool using which the system can improve its cooperative nature and inferential capabilities. *Id.*, 890-891.

A POSITA would have been motivated to incorporate the plan-based methodology and discourse analysis teachings of Huang into the system of Kennewick. Ex. 1002, ¶¶135-138. The two references share common goals of improving natural responses to user utterances based on contextual information. Ex. 1003, [0011] (“robustness to partial failure promotes the feeling of a natural response”); Ex. 1011, 892 (plan-based system permits “digression, since the user’s intention model has been built into the plan”). And a POSITA would also have sought out the improved inferential ability resulting from the plan approach in

furtherance of Kennewick’s disclosed goal to “minimize the interaction with the user required to deliver a response.” Ex. 1003, [0032]; Ex. 1002, ¶138. A POSITA would have been able to use Kennewick’s dialog history to derive a conversational goal, and a plan, following the teachings of Huang. Ex. 1002, ¶138. By considering the intent and aims of the user, and the roles of the system and user in terms of information allocation, Huang’s system provides additional context that would add another layer of information to Kennewick’s dialog history.

Accordingly, a POSITA would have found it obvious to modify the teachings of Kennewick, enhancing the disclosed system with the plan-based conversational adaptation techniques of Huang. *Id.*, ¶¶135-139.

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

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

Huang discloses using conversational goals, roles, and corresponding information allocation to determine meaning. The conversational “goals of listener and speaker in the interaction,” also known as intent, form one of four main areas “from which meaning arises” in spoken language understanding. Ex. 1011, 854. The other three areas—context, content, and assumptions—relate to the roles of and information allocated among the participants in the conversation. *Id.*

In determining meaning based on goals, Huang describes modeling “goals related to both the task (*Show me all flights ...*), and the dialog process itself (*Please repeat ...*) of the users.” *Id.*, 859 (emphasis in original). Huang also discloses organizing a dialog into a hierarchy of goals, including a scenario (desired output of the exchange), goals (of the user and the system), and sub-goals (details within the goals). *Id.*, 893-894, Fig. 17.22. Only by understanding the user’s intentions—or goals—can the system “respond appropriately.” *Id.*, 862.

Huang teaches that the roles of conversational participants affect the “system’s view of how the dialog should proceed.” *Id.*, 866. System and user roles are on a continuum; on one end the system “directs” the user, on the other end the system “accepts” the user, and in between the system “guides” or “inform[s]” the user. *Id.* Huang explains that to improve human-computer dialog, such varied roles allow a natural interaction:

It is often more natural, however, to allow for *mixed initiative systems*, where interaction starts with a user’s query or command and the system attempts to derive, via inference or further questioning of the user, all information needed to understand and process a complete transaction. When the user knows clearly what he wants and the system has no trouble catching up, the user is in the driver’s seat. However, when the system detects that the user is in a state of confusion, or when it has trouble getting user’s intention, the machine will offer guidance or negotiate with the user to steer the dialog back on track.

Id., 860 (emphasis in original).

A POSITA would understand that these varied roles also correspond to a continuum in information allocation between the participants in the conversation. Ex. 1002, ¶¶140-143. For example, in a system-directed dialog, the system “retains complete dialog control” and “initiates any dialog necessary to obtain completion of information from the user”; the system has more information allocation. *Id.* On the opposite end of the spectrum, with the system simply accepting user commands, the user has more information allocation.

Huang describes using these dialog characteristics, among others, to determine the meaning of a user utterance and a corresponding “plan” for responding to and meeting the user’s goals. Ex. 1011, 888-889. A POSITA would have been motivated to use this approach, which “permit[s] a more flexible mode of interaction,” *id.*, 889, 892, in the system of Kennewick, which seeks to provide a “natural environment” for users. Ex. 1003, Abstract; Ex. 1002, ¶144.

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

As described in the previous section, Huang discloses using goal, roles, and information allocation characteristics of a dialog to establish the intended meaning of an utterance. Huang captures these characteristics in a “plan” or “model of all

[dialog] participants’ goals, commitments, and beliefs,” according to which the system interprets user intent. Ex. 1011, 889. The “plan” concept thus meets a POSITA’s understanding of “conversation type” in view of the ’681 patent, which identifies three types based on the same goal, role, and information allocation dialog characteristics: query (getting information from a supporter), didactic (getting information from a teacher), and exploratory (sharing information). Ex. 1001, 15:8-34. Ex. 1002, ¶145.

The plan-based approach to dialog requires observing the acts (utterances) of the user, inferring the user’s plan, identifying obstacles, setting a goal of overcoming the obstacles, and planning to achieve the goal. Ex. 1011, 890. The system uses this plan to make inferences based on the “beliefs, desires, and intentions” of the user that satisfy the goal. *Id.*, 891. Additionally, if the system determines a conflict between its model (plan) and the “utterances and speech acts” of the user during a dialog, it can change its plan. *Id.* In other words, the system continually determines a plan consistent with the progress of a dialog and uses that plan to identify the intended meaning behind each utterance. *Id.*, 892 (“the user’s intention model has been built into the plan”). Again, this is consistent with how a POSITA would understand “conversation type” based on the ’681 patent’s disclosure that a “conversation type may change dynamically.” Ex. 1001, 5:43-47; Ex. 1002, ¶146.

Due to the similar goals of the plan-based approach disclosed by Huang and the system of Kennewick, also described in the previous section, a POSITA would have been motivated to modify Kennewick to use the plans of Huang in modeling dialog participants' goals, commitments, and beliefs. Ex. 1002, ¶147. The modified system would use a plan model as part of establishing the intended meaning of a user's utterance. *Id.* As Huang discloses, the plan-based approach may “overcome the rigidity and shallowness of dialog grammars and templates.” Ex. 1011, 888. It can “permit a more flexible mode of interaction.” *Id.*, 889. This aligns well with the goals of Kennewick to avoid presenting information in a “rigid, highly formatted, or structured manner” and to “create a natural question and response environment.” Ex. 1003, [0024], [0032]. Thus, a POSITA would have been motivated to modify the system of Kennewick by adopting the concept of plans as taught by Huang and using them to improve responses to the user. Ex. 1002, ¶¶145-147.

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

As described above, Kennewick discloses grammatically or syntactically adapting the response based on an intended meaning (Section IX.A.1[1.g]) and Huang discloses identifying a conversation type (Section IX.B.1(claim 5)). The conversation type, or “plan” of Huang, models the “user’s intention.” Ex. 1011, 892. Huang further discloses that the “cooperative process” to satisfy the goal of each

plan may require the system to make “the minimal queries need to verify and choose among alternative hypotheses” or base its response on “pure inference, with perhaps a confirmation step.” *Id.*, 891. Additionally, “a cooperative system” can help clarify the user’s input. *Id.*, 892. For example, Huang discloses two potential responses to a user’s utterance, “I would like to fly to Boston.” *Id.*, 893. The first, an explicit confirmation, asks “Do you want to fly to Boston?” *Id.* The second is implicit: “When do you want to fly to Boston?” *Id.* The system selects the response based on a confidence measure, which a POSITA would understand should be determined based on the system’s understanding of conversational participants’ goals, commitments, and beliefs—the plan, or conversation type. Ex. 1002, ¶148.

Huang teaches the desirability of submitting “minimal queries” to achieve the goal, and Kennewick similarly discloses a goal of “minimiz[ing] the interaction with the user required to deliver a response.” Based on these similar goals, and to take advantage of the benefits of a plan-based approach as disclosed by Huang, *see* Section IX.B.1(claim 5), a POSITA would have been motivated to modify the system of Kennewick, in view of Huang, to use conversation type (plan) as a factor in adapting responses to the user. Ex. 1002, ¶149.

3. Claim 17

The additional limitations in dependent Claim 17 are substantially similar to those of claim 5 other than minor variations in syntax/tense; thus Kennewick and

Huang render claim 17 obvious for the reasons described in Section IX.B.1, with limitations 17.a-17.b corresponding to limitations 5.a-5.b, respectively. Ex. 1002, ¶¶150-151.

4. Claim 20

The additional limitation in this dependent claim is substantially similar to that of claim 8, other than rephrasing “the conversation type” as “conversation type,” and is obvious for the reasons in Section IX.B.2. Ex. 1002, ¶152.

5. Claim 29

The additional limitations in dependent claim 29 are substantially similar to those of claim 5 other than minor variations in syntax/tense and the added recitations of “processors configured to” perform the claimed functions. As discussed regarding parent claim 25 in Section IX.A.7[25.b], Kennewick discloses processors to perform the system functions. Thus, Kennewick and Huang render claim 29 obvious for the reasons described in Section IX.B.1, with limitations 29.a-29.b corresponding to limitations 5.a-5.b, respectively. Ex. 1002, ¶¶153-154.

6. Claim 32

The additional limitation in this dependent claim is substantially similar to that of claim 8, other than rephrasing “further grammatically or syntactically adapts the response” as “generate the grammatically syntactically adapted response”, and is obvious for the reasons in Section IX.B.2. Ex. 1002, ¶155.

C. Ground 3: Claims 10, 22, 34, 38, 40, and 42 are rendered obvious by Kennewick in view of Seneff.

As described in Section IX.A, Kennewick renders obvious independent claims 1, 13, 25, 37, 39, and 41, from which claims 10, 22, 34, 38, 40, and 42 depend. This section identifies additional disclosures of Kennewick and Seneff that render these dependent claim limitations obvious. Ex. 1002, ¶156.

As described in Section VIII, a POSITA would have considered prior art references in shared fields and could have combined their teachings in a modular way. Kennewick and Seneff each describe spoken dialog systems. Ex. 1003, Abstract (“complete speech-based natural language query and response environment”); Ex. 1025, §1 (describing complex, multi-variable dialog system for making flight reservations). Both systems seek to provide accurate responses to spoken user input. Kennewick discloses a multi-stage process that resolves ambiguities and inconsistencies based on a wealth of contextual information. Ex. 1003, [0028]-[0029]. Seneff describes a dialog strategy that involves a hypothesis selection and verification process. Ex. 1025, §1.

The hypothesis selection process of Seneff first identifies a set of “alternative candidates for the meaning of the utterance,” and then “singles out the most promising” one based on a confidence scoring process. Ex. 1025, §2.2. The scoring process considers several factors, including in-focus semantic categories or preferential treatment for or against certain words. *Id.* Once the scoring is complete,

the system passes the winning candidate to the dialog manager for evaluation relative to the context of the dialog. *Id.*

Kennewick does not explicitly disclose scoring hypotheses for the *meaning* of an utterance. However, it does teach the use of similar scoring methodologies to evaluate the most likely context or domain for a user's utterance, which is used to determine its meaning. Ex. 1003, [0161], [0170]-[0171]. A POSITA would have been motivated to search for other stages of Kennewick's speech recognition and response process that could be improved by scoring, such as in determining the meaning of an utterance. Ex. 1002, ¶¶157-159. It would have been obvious to modify the system of Kennewick to use the meaning hypothesis selection process of Seneff in furtherance of Kennewick's goal to provide reliable, expected results to a user's questions or commands. Ex. 1003, Abstract; Ex. 1002, ¶159. Additionally, the context-rich system of Kennewick would have had the information needed to support the evaluation of hypotheses as taught by Seneff. Ex. 1002, ¶159. Accordingly, a POSITA would have found it obvious to modify the teachings of Kennewick, enhancing the system with Seneff's confidence-scoring system for determining the meaning of an utterance. *Id.*, ¶¶157-159.

1. Claim 10: “The method of claim 1, further comprising:”

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

Seneff discloses a hypothesis selection process for understanding a user utterance that generates multiple preliminary hypotheses of what the intended

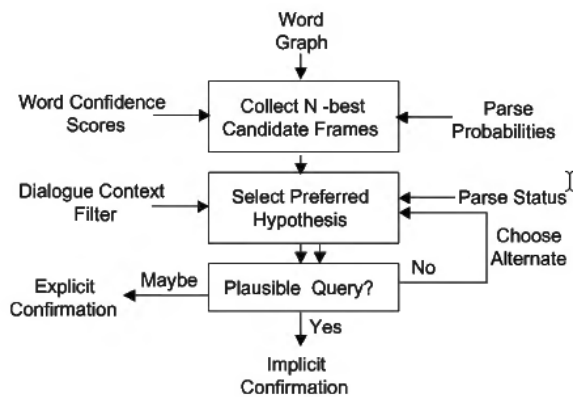


Figure 4: A block diagram of the process of hypothesis selection and verification.

meaning of the utterance may be.

Ex. 1025, §2.2, Fig. 4, left. The goal of the process is to identify the “most plausible solution” and determine “what the user has plausibly intended” in each utterance. *Id.*, §7.

The process uses “relevant information [that] is retained from prior turns” and “maintained and updated continually throughout the user’s conversation.” Ex. 1025, §§1, 4. The first step creates a “word graph, representing multiple sentence hypotheses, with associated confidence scores for each word in the graph.” *Id.*, §2.2. Next, the TINA natural language system parses the word graph into “candidate semantic frames”—multiple hypotheses—each one labeled with a parse rating indicating how many words in each hypothesis the parse tree accounts for. *Id.*

The process then uses a “simple heuristic” to identify the “most promising candidate” by choosing “the highest scoring full-parse theory.” *Id.* Thus, Seneff discloses an initial interpretation that has a highest confidence based on highest score. Seneff also describes using any existing dialog “context conditions” to select an initial interpretation that may “preferentially favor an otherwise sub-optimal theory,” such as one that lacks the highest raw score according to the heuristic. *Id.* In this scenario, the interpretation with the highest confidence level may be one that does not have the highest raw score. *Id.*; Ex. 1002, ¶162. Regardless of the score, a POSITA would understand Seneff’s disclosure of “singl[ing] out” a “most promising hypothesis” to indicate that there is at least one interpretation the system deems to have a highest confidence level. Ex. 1025, §2.2; Ex. 1002, ¶162.

The purpose of the process to assure “that the system underst[ands] correctly each user query” in Seneff aligns with the goal of Kennewick to provide appropriate results even where “the question asked or the responses received are incomplete, ambiguous or subjective.” Ex. 1025, §1; Ex. 1003, [0002]. Thus, a POSITA would have been motivated to modify the system of Kennewick to include the hypothesis selection process taught by Seneff and improve the system’s ability to understand user utterances accurately. Ex. 1002, ¶163. In the modified system, the selection process would rely on the wealth of contextual information disclosed in Kennewick

to score and evaluate recognition hypotheses as taught by Seneff, resulting in an initial interpretation that has a highest confidence level. *Id.*, ¶¶160-163.

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

As Seneff discloses, the initial selection of the candidate interpretation representing the “highest-scoring full-parse theory” may be incorrect based on contextual conditions established during the current dialog. Ex. 1025, §2.2. In this scenario, the hypothesis selection process may “[a]sk the user for explicit confirmation” or “[i]nitiate a subdialogue asking for confirmation.” *Id.*, §§2.2, 3 (describing exemplary confirmation dialogs). In the combined system of Kennewick and Seneff, as described in the preceding section, the initial interpretation would thus make it into the dialog history, *i.e.*, the short-term shared knowledge. If the user responds that the initial interpretation is wrong, the system must ensure that the incorrect interpretation does not cause further errors down the line. Accordingly, a POSITA would have found it obvious to try one of two options: (1) remove the interpretation from the dialog history, or (2) to flag the interpretation as erroneous, effectively removing it from the contextual information used to evaluate future utterances. Ex. 1002, ¶164.

A POSITA would not have left the incorrect interpretation unaddressed, which would run counter to the goals of both Kennewick and Seneff to improve the understanding capabilities of the system. Ex. 1025, §1; Ex. 1003, [0002]; Ex. 1002, ¶165. Additionally, Seneff identifies a problem specific to cities “not in the recognizer’s vocabulary” where the system repeatedly substitutes the “same incorrect hypothesis” into the system’s understanding of an utterance. Ex. 1025, §4. Retaining incorrect interpretations as-is in the dialog history, even after the system has already confirmed them as wrong, would exacerbate this problem and could cause repeated incorrect hypothesis substitutions in situations beyond the specific vocabulary issue described. Ex. 1002, ¶165.

Upon determining that the initially selected interpretation is wrong, the system may choose an “alternative hypothesis from the *N*-best list, that may be more appropriate pragmatically.” Ex. 1025, §2.2; *see also* §4 (maintaining frames of “alternative recognizer hypotheses of dates, for possible later reselection”). As the *N*-best list provides an ordered set of alternatives, a POSITA would have known that the system would choose the alternative hypothesis with the next-highest confidence level based on available information, including scores, ratings, and any further confirmation or sub-dialog with the user. Ex. 1002, ¶166.

For the reasons above, the combined system of Kennewick and Seneff thus renders this limitation obvious. *Id.*, ¶¶164-167.

2. Claim 22

The additional limitations in dependent claim 22 are substantially similar to those of claim 10 other than minor variations in syntax/tense, broader in that 22.a omits that the speech recognition engine is “coupled to the voice input device and the conversational speech engine,” and adding in 22.b that the speech engine “identifies the context associated with the utterance.” As discussed regarding parent claim 13 in Section IX.A.4[13.d], the speech engine identifies the context. Thus, Kennewick and Seneff render claim 29 obvious for the reasons described in Section IX.C.1, with limitations 22.a-22.b corresponding to limitations 10.a-10.b, respectively. Ex. 1002, ¶¶168-169.

3. Claim 34

The additional limitations in dependent claim 34 are substantially similar to those of claim 10 other than minor variations in syntax/tense. 34.a is broader than 10.a as it omits that the speech recognition engine is “coupled to the voice input device and the conversational speech engine.” 34.b substitutes “one or more processors are further configured to” for “conversational speech engine” in 10.b and additionally recites “identify the context associated with the utterance.” As discussed regarding parent claim 25 in Sections IX.A.7[25.b], [25.e], Kennewick discloses processors to perform the system functions including identifying context. Thus, Kennewick and Seneff render claim 34 obvious for the reasons described in

Section IX.C.1, with limitations 34.a-34.b corresponding to limitations 10.a-10.b, respectively. Ex. 1002, ¶¶170-171.

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

As described in Section IX.C.1[10.a], the hypothesis selection process of Seneff, designed to make sure the “system understood correctly each user query,” involves various indications of certainty that the hypotheses truly capture the user’s intent. Ex. 1025, §§1, 2.2. Each candidate hypothesis has an “overall sentence score” obtained by combining “confidence scores” of the words with “acoustic and linguistic scores.” *Id.*, §2.2. Each hypothesis also has a “parse status” label indicating how much of the hypothesis is accounted for in a single parse tree (and how much is ignored). *Id.* The process selects the “most promising” hypothesis based on the “highest scoring” theory with the highest parse status, in view of “context conditions” based on the dialog. *Id.* These conditions, including in-focus “semantic categories” or “favored”/“disfavored” words, help identify which hypothesis is the most appropriate in the context of the dialog. *Id.* Finally, a “dialogue manager” evaluates whether the selected hypothesis is consistent with prior dialog. *Id.* This last step further verifies whether the hypothesis comports with the user’s intended meaning. Ex. 1002, ¶172. Seneff’s approach thus discloses

selecting a hypothesis based on a degree of certainty regarding the user’s intent in speaking an utterance.

The hypothesis may be “implicit,” requiring further processing to fill in information lacking in the utterance. Section VII.A (construction of “implicit hypothesis”); Ex. 1002, ¶173. For example, the dialog manager in Seneff infers certain kinds of information, such as to resolve phrases like “the following Friday” or “three days later” into explicit dates. Ex. 1025, §2.3.

Kennewick similarly discloses making inferences and substituting portions of the utterance with corresponding values relevant to the context of a dialog. Ex. 1003, [0164]. In addition to the reasons described in Section IX.C.1[10.a], a POSITA would have been motivated to combine Kennewick with Seneff for this reason too. Ex. 1002, ¶174. The resulting system renders this claim obvious.

5. Claim 40

The additional limitation of this dependent claim is identical to that of claim 38, and is thus obvious for the reasons in Section IX.C.4. Ex. 1002, ¶175.

6. Claim 42

The additional limitation of this dependent claim is identical to that of claim 38, and is thus obvious for the reasons in Section IX.C.4. Ex. 1002, ¶176.

D. Ground 4: Claims 6-7, 18-19, and 30-31 are rendered obvious by Kennewick and Huang, further in view of Seneff.

As described in Section IX.B, Kennewick and Huang render claims 5, 17, and 29 obvious, from which claims 6-7, 18-19, and 30-31 depend. This section identifies additional disclosures of Seneff that render these dependent claim limitations obvious. Ex. 1002, ¶177.

As described in Sections IX.B and IX.C, a POSITA would have combined Kennewick with Huang, and would also have combined Kennewick with Seneff. A POSITA would also have been motivated to modify the combined Kennewick and Huang system to include the hypothesis selection process of Seneff described above in Section IX.C. Ex. 1002, ¶178. The modification of Kennewick to include the plan-based methodology and discourse analysis teachings of Huang would provide additional contextual information for use during the hypothesis selection process. Ex. 1011, 881 (“current dialog status” and “discourse knowledge” can be important for correct interpretations), 892 (plan-based systems enable tracking “*dialog focus*, or temporary centers of attention”) (emphasis in original); Ex. 1025, §2.2 (hypothesis selection may rely on “context conditions” including “semantic categories that are in focus”). A POSITA would have expected the additional information to improve the ability of the system to understand user utterances, as taught by Seneff. Ex. 1025, §1 (for natural dialog interfaces, “it becomes important

to draw on as many constraints as possible to aid in the hypothesis selection task”); Ex. 1002, ¶178.

1. Claim 6

The additional limitation of this dependent claim is substantially similar to that of claim 38, but is broader in that it recites a “hypothesis” rather than an “implicit hypothesis,” and omits “corresponding” from the “degree of certainty.” It also specifies that there are “one or more words in” the utterance, which does not affect the applicability of the prior art. This limitation is thus obvious for the reasons in Section IX.C.4. Ex. 1002, ¶179.

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

As described regarding claim 10, Seneff discloses a way of improving the ability of a spoken dialog system to understand the meaning behind a user’s utterance through a hypothesis selection process. Ex. 1025, §2.2; Section IX.C.1. Seneff further teaches determining a degree of certainty for a hypothesis based on confidence measures and contextual information, as described regarding claim 38 above. Section IX.C.4.

A POSITA would have been motivated to modify the combined system of Kennewick and Huang to further incorporate the hypothesis selection process of Seneff and improve its ability to understand user utterances, as explained above with respect to this ground. Ex. 1002, ¶¶180-181. Because the combined system would assess the degree of certainty for the preliminary interpretation based on at least two of the three categories of information recited in this claim (“confidence level” and “information associated with the identified context”), the claim is obvious. *See Ex parte Kubin*, 2007 Pat. App. LEXIS 13, at *10 (B.P.A.I. May 31, 2007), *aff’d*, 561 F.3d 1351, 1361 (Fed. Cir. 2009) (single obvious species renders claimed genus unpatentable).

3. Claim 18

The additional limitation in this dependent claim is identical to that of claim 6, and is obvious for the reasons in Section IX.D.1. Ex. 1002, ¶182.

4. Claim 19

The additional limitation in this dependent claim is substantially similar to that of claim 7 other than rephrasing “generating” as “generate,” and is broader in that it omits the speech recognition engine being “coupled to the voice input device and the conversational speech engine.” Thus, this claim is obvious for the reasons in Section IX.D.2. Ex. 1002, ¶183.

5. Claim 30

The additional limitation of this dependent claim is identical to that of claim 6 and is obvious for the reasons in Section IX.D.1. Ex. 1002, ¶184.

6. Claim 31

The additional limitation in this dependent claim is substantially similar to that of claim 7 other than rephrasing “generating” as “generate” and replacing “conversational speech engine” with “one or more processors are further configured to.” It is also broader in that it omits the speech recognition engine being “coupled to the voice input device and the conversational speech engine.” As described regarding top-level parent claim 25, Kennewick discloses processors performing the speech-related functions of the system. Thus, this limitation is obvious for the reasons in Sections IX.A.7[25.b] and IX.D.2. Ex. 1002, ¶185.

E. Ground 5: Claims 11-12, 23-24, and 35-36 are rendered obvious by Kennewick in view of Coffman.

As described in Section IX.A, Kennewick renders obvious independent claims 1,13, and 25, from which claims 11-12, 23-24, and 35-36 depend. This section identifies additional disclosures of Coffman that render these dependent claim limitations obvious. Ex. 1002, ¶186.

As described in Section VIII, a POSITA would have considered prior art references in shared fields and could have combined their teachings in a modular way. Kennewick and Coffman both describe similar systems for providing natural,

conversational user experiences. Ex. 1003, [0002]; Ex. 1027, 1:9-14. The Kennewick system “makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment.” Ex. 1003, [0011]. In Coffman, the conversational application similarly supports “free flow interactions (e.g., mixed initiative dialogs) ... preferably using short term and long term context (including previous input and output) to disambiguate and understand the user’s intention.” Ex. 1027, 4:9-16, 6:12-38. The system also applies mechanisms like Kennewick’s for disambiguating user input including, for example, context resolution based on the current conversation and user history. *Id.*, 9:25-33.

Additionally, Kennewick discloses a natural language interface with which a user may interact through multiple modalities, including “live audio, live video, electronic messaging or email, instant messaging, or other techniques.” Ex. 1003, [0065], [0076], [0125]. However, Kennewick’s disclosure lacks details about managing the modalities of the interface. Coffman provides additional information regarding the management of “multi-modal input/output,” emphasizing the importance of multi-modal interaction to address the need “to increase the bandwidth of man-machine communication.” Ex. 1027, 1:36-48, 3:14-15. Accordingly, a POSITA would have found it obvious to modify the Kennewick system to adopt the multi-modal interaction mechanisms taught by Coffman,

enhancing the ability of the system to address information in additional forms.

Ex. 1002, ¶¶187-188.

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

The Coffman system supports interaction through modalities “[s]uch as speech (e.g., authored in VoiceXML), visual (GUI) (e.g., HTML (hypertext markup language)), constrained GUI (e.g., WML (wireless markup language), CHTML (compact HTML), HDML (handheld device markup language)), and a *combination of such modalities (e.g., speech and GUI).*” Ex. 1027, 4:17-22 (emphasis added). Coffman also teaches that “each modality (or combination of modalities) may be implemented as a full NL (natural language) user interface.” *Id.*, 4:22-26. Thus, a POSITA would have understood that Coffman discloses a variety of non-voice inputs the user may provide in combination with speech. Ex. 1002, ¶189. As explained above, it would have been obvious for a POSITA to modify the system of Kennewick, which already accepts input in multiple modalities, Ex. 1003, [0065], to use the teachings of Coffman and accept combined voice and non-voice inputs. Ex. 1002, ¶189. This modification would provide additional flexibility to the user and serve “to increase the bandwidth of man-machine communication” as Coffman explains. Ex. 1027, 1:36-48.

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

The Coffman system provides multi-modal output using an I/O manager which “generates the message in one or more modalities” and then “task managers to drive the output generation to generate an abstract output event.” Ex. 1027, 3:14-19. The I/O manager converts the output event “into one or more modalities for presentation to the user.” *Id.*, 3:19-21. For the actual presentation, the system uses modality agents. *Id.*, 21:60-22:8, *see also* Figs. 8, 9, below.

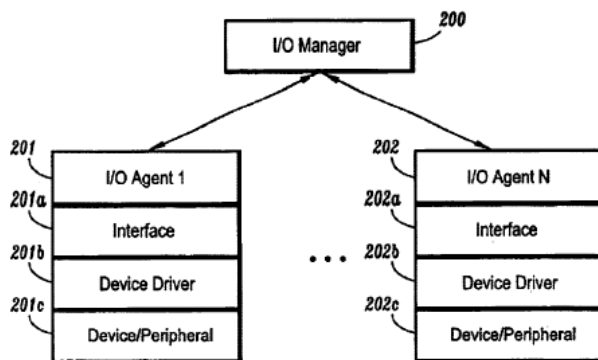


FIG. 8

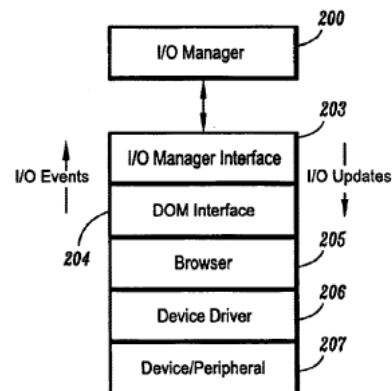


FIG. 9

Each agent “handles a particular modality,” acting as an interface to the respective device driver “(e.g. Accessibility API for GUI, DOM for a browser, telephony API for telephone speech, etc.).” *Id.*, 22:5-14. Through these agents, the I/O manager submits “output events” for execution by “one or more different channels (device, modality etc.).” *Id.*, 21:52-59. An agent may request additional

information “specifying how the output data should be provided.” *Id.*, 22:44-46. The system may synchronize “between events from multiple modalities,” permitting “the use of multiple modalities in a complementary fashion.” *Id.*, 21:26-28, 22:57-59. Thus, Coffman discloses a multi-modal output that includes one or more non-voice outputs, for example “speech and GUI.” *Id.*, 4:22.

As explained above, it would have been obvious for a POSITA to modify the system of Kennewick, which also discloses various output modalities, to use the teachings of Coffman in providing a “speech and GUI” multi-modal output. Ex. 1003, [0125] (using a GUI as a “complement to the speech interface”), [0145] (“email, the text to speech engine, a message to a pager, or a text and graphics display”). As with multi-modal inputs in the previous section, this modification would provide additional flexibility to the user and serve “to increase the bandwidth of man-machine communication” as Coffman explains. Ex. 1027, 1:36-48; Ex 1002, ¶¶190-192. This modified system alone renders the claim obvious, as it discloses one of the two types of content recited for the claimed “multi-modal output.” See *Ex parte Kubin*, 2007 Pat. App. LEXIS 13, at *10 (single obvious species renders claimed genus unpatentable); Ex. 1002, ¶193.

3. Claim 23

The additional limitation of this dependent claim is identical to that of claim 11, and is obvious for the reasons in Section IX.E.1. Ex. 1002, ¶194.

4. Claim 24

The additional limitation of this dependent claim is identical to that of claim 12, and is obvious for the reasons in Section IX.E.2. Ex. 1002, ¶195.

5. Claim 35

The additional limitation of this dependent claim is identical to that of claim 11, and is obvious for the reasons in Section IX.E.1. Ex. 1002, ¶196.

6. Claim 36

This limitation is substantially similar to that of claim 12, but instead of specifying that the “conversational speech engine generates the response in a multi-modal output,” it recites that the “grammatically or syntactically adapted response” of parent claim 25 comprises a “multi-modal output.” This difference does not affect the applicability of the prior art. Thus, the additional limitation of this dependent claim is obvious for the reasons in Section IX.E.2. Ex. 1002, ¶197.

F. Ground 6: Claims 3-4, 15-16, and 27-28 are rendered obvious by Kennewick in view of Mitsuyoshi

As described in Section IX.A, Kennewick renders obvious claims 2, 14, and 26, from which claims 3-4, 15-16, and 27-28 depend. This section identifies additional disclosures of Mitsuyoshi that render these dependent claim limitations obvious. Ex. 1002, ¶198.

As described in Section VIII, a POSITA would have considered prior art references in shared fields and could have combined their teachings in a modular

way. A POSITA would have been motivated to combine Kennewick and Mitsuyoshi, both of which describe systems for processing natural language voice input. Ex. 1003, [0002]; Ex. 1028, [0066]. Kennewick teaches storing and using a wide array of contextual data to interpret and respond to users, including both dialog history (short-term) and user profile (long-term) data. Ex. 1003, [0029], [0031], [0160]. Mitsuyoshi similarly discloses a knowledge database “for storing situation information showing a past situation, a past episode, and a result of the past situation and episode” and updating the database with a “newly inputted situation and a result of the new situation.” Ex. 1028, [0060]. Kennewick also describes using “personalities” simulating the “behavioral characteristics of real humans” to make responses more natural. Ex. 1003, [0035]. This aligns with Mitsuyoshi’s goal of “outputting sensibility more akin to that of a human.” Ex. 1028, [0016].

Mitsuyoshi addresses this goal by detecting and managing emotion data based on voice input. *Id.*, [0019]. Mitsuyoshi teaches several ways to manage the data, including the use of an “oblivion processing unit” that deletes emotion data stored for a certain period of time, but excludes larger emotional shifts which “similarly to a memory of a human,” “may be useful in the future” even if it is old. *Id.*, [0028]. This preserves certain data more likely to be useful in a way similar to how human memory works. *Id.*, [0028], [0125]-[0126]; Ex. 1002, ¶200.

At the time of the '681 patent, emotion was a well-known component of short-term conversation knowledge that could be considered in a speech processing system. Ex. 1002, ¶201. A POSITA would have found it obvious that methods of managing one component of conversational data would be applicable to other kinds of components too. *Id.* Thus, a POSITA would have modified the system of Kennewick to adopt the methods of managing emotion data disclosed by Mitsuyoshi for managing other aspects of the short-term knowledge disclosed by Kennewick. *Id.*, ¶¶199-201. The resulting system would achieve the benefits of conserving storage capacity and reducing processing speed by remove data less likely to be useful. Ex. 1028, [0062]-[0063].

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

The Kennewick system discloses “age of the context” as a factor in weighting information on the context stack, which stores part of the dialog history (short-term knowledge). Ex. 1003, [0170]. Mitsuyoshi uses the same concept to evaluate and cull stored emotion data to reclaim capacity and improve processing speed. Ex. 1028, [0062]-[0063], [0126]. The “oblivion processing unit” of Mitsuyoshi chooses to preserve data when (1) a certain period of time has not yet elapsed, so the data is too new (*id.*, [0129]); (2) the data shows a significant change in emotion (*id.*,

[0130]); (3) the emotion and speech data match a “pattern representing a ‘strongly impressive’ state” (*id.*, [0131]); or (4) the speech data matches predetermined content for “significantly impressive speech” (*id.*, [0132]). The patterns and predetermined content are stored on a database that changes based on learned information, resulting in different outcomes over time. *Id.*, [0134]. Thus, Mitsuyoshi describes a test for whether to cull older data based on psychological characteristics, taking elapsed time into account. *Id.*, [0228] (“degree of oblivion and the priorities of the respective data can be changed according to the individuality”); Ex. 1002, ¶202.

Since one goal of Kennewick is “to maintain a natural interaction” such that “responses need to be returned rapidly to the user,” it would have been obvious to modify Kennewick to manage short-term information in the way taught by Mitsuyoshi, which emulates how human memory works, saves on storage, and avoids lowering processing speed, while preserving information more likely to be useful. Ex. 1003, Abstract, [0008]; Ex. 1028, [0062]-[0063]; Ex. 1002, ¶203. This would also make way for more recent and timely information in the short-term context stack, another reason to expire short-term information after the sufficient passage of time. Ex. 1002, ¶203.

The modified system would decide whether to cull information from the short-term context stack based on the passage of time and the equivalent of “pattern” or

“predetermined content” criteria representing the significance of each piece of information. *Id.*, ¶204. The system could base these criteria on user preferences or other historical information Kennewick already tracks in its user profile. Like the pattern database of Mitsuyoshi, Kennewick’s user profile also automatically updates and would result in the evaluation changing over time as the user continues using the system. Ex. 1028, [0134]; Ex. 1003, [0031]; Ex. 1002, ¶204.

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

Expiring short-term information, which the combined system of Kennewick and Mitsuyoshi discloses, Section IX.F.1, could be accomplished in several obvious ways: simple deletion or retention in another form or location. Ex. 1002, ¶205. A POSITA would have found it obvious to retain relevant, though expired, information for later reference. *Id.* Data failing to achieve the degree of significance required under the Mitsuyoshi test to be preserved as part of the short-term context could still be useful as part of the long-term context, including as a “history of questions asked,” past interests of the user, or previous transactions conducted. Ex. 1003, [0031]. Thus, the combination of Kennewick and Mitsuyoshi renders this claim obvious.

3. Claim 15

The additional limitation of this dependent claim is substantially similar to that of claim 3, other than replacing “accumulating” with “to accumulate” and replacing “includes expiring” with “the computer-executable instructions are further operable when executed to expire.” As described regarding top-level parent claim 13, Kennewick discloses computer-executable instructions for performing the functions of the system. Thus, this limitation is obvious for the reasons in Sections IX.A.4[13.pre] and IX.F.1. Ex. 1002, ¶206.

4. Claim 16

The additional limitation of this dependent claim is substantially similar to that of claim 4, other than replacing “accumulating” with “to accumulate” and replacing “includes updating” with “the computer-executable instructions are further operable when executed to update.” As described regarding top-level parent claim 13, Kennewick discloses computer-executable instructions for performing the functions of the system. Thus, this limitation is obvious for the reasons in Sections IX.A.4[13.pre] and IX.F.2. Ex. 1002, ¶207.

5. Claim 27

The additional limitation of this dependent claim is substantially similar to that of claim 3, other than replacing “accumulating” with “to accumulate” and replacing “includes expiring” with “the one or more processors are further configured to expire.” As described regarding top-level parent claim 25, Kennewick

discloses processors performing the speech-related functions of the system. Thus, this limitation is obvious for the reasons in Sections IX.A.7[25.b] and IX.F.1. Ex. 1002, ¶208.

6. Claim 28

The additional limitation of this dependent claim is substantially similar to that of claim 4, other than replacing “accumulating” with “to accumulate” and replacing “includes updating” with “the one or more processors are further configured to update.” As described regarding top-level parent claim 25, Kennewick discloses processors performing the speech-related functions of the system. Thus, this limitation is obvious for the reasons in Sections IX.A.7[25.b] and IX.F.2. Ex. 1002, ¶209.

X. THE PETITION SHOULD NOT BE DENIED UNDER §§ 314 OR 325

Sections 314(a) and 325(d) are not grounds for denial of this petition based on the disclosure of Kennewick on an IDS during prosecution of the '681 patent. The PTAB considers six factors to determine whether it should exercise discretion under § 325(d). *Becton, Dickinson & Co. v. B. Braun Melsungen AG*, IPR2017-01586, Paper No. 8 at 17-18 (PTAB Dec. 15, 2017) (precedential). Factors 3, 5, and 6 relate to the examiner’s evaluation of the reference during prosecution. *Id.* These factors do not apply. Kennewick was cited in an IDS that identified over 120 references, is among over 370 references disclosed during

prosecution, and was not substantively applied by the examiner. Ex. 1008, 72 (listing Kennewick on a September 27, 2007 IDS of five pages of references); Ex. 1001, 3 (listing Kennewick in nearly six columns of references). “[T]he Board has consistently declined exercising its discretion under Section 325(d) when the only fact a Patent Owner can point to is that a reference was disclosed to the Examiner during the prosecution.” *Amgen Inc. v. Alexion Pharm., Inc.*, IPR2019-00739, 2019 WL 4132683, at *25-26 (PTAB Aug. 30, 2019); *Amazon.com, Inc., v. M2M Solutions LLC*, IPR2019-01204, Paper No. 14 (PTAB Jan. 23, 2020) (citations omitted) (refusing to exercise discretion under 325(d) where the examiner did not substantively consider the cited material, which was among “a few hundred” identified across multiple IDSs). Additionally, factors 1, 2 and 4 relate to the similarity between the reference and the prior art substantively evaluated by the Examiner during prosecution. *Becton*, Paper 8 at 17–18. These factors also do not apply because Kennewick is materially different from the prior art that the Examiner evaluated. Indeed, Kennewick discloses the subject matter that the Examiner identified as missing from the prior art. Ex. 1008, 1042 (Notice of Allowance). Accordingly, the *Becton* factors do not favor denial of this petition under § 325(d).

XI. CONCLUSION

For the reasons given above, the challenged claims are unpatentable. Ex. 1002, ¶210. Accordingly, Petitioners respectfully request the institution of *inter partes* review on all claims of the '681 patent.

Dated: July 28, 2020

Respectfully submitted,

FENWICK & WEST LLP

/J. David Hadden/

J. David Hadden

Reg. No. 40,629

Attorneys for Petitioners Amazon.com, Inc.,
Amazon.com LLC, Amazon Web Services,
Inc., A2Z Development Center, Inc. d/b/a
Lab126, Rawles LLC, AMZN Mobile LLC,
AMZN Mobile 2 LLC, Amazon.com Services,
Inc. f/k/a Amazon Fulfillment Services, Inc.,
and Amazon.com Services LLC (formerly
Amazon Digital Services LLC)

**CERTIFICATE OF COMPLIANCE
WITH TYPE-VOLUME LIMITATION**

Pursuant to 37 CFR §42.24 et seq., the undersigned certifies that this Petition complies with the 14,000-word type-volume limitation. This Petition contains 13,832 words, excluding the parts of the Petition exempted from the type-volume limitation.

Dated: July 28, 2020

FENWICK & WEST LLP

/J. David Hadden/

J. David Hadden

Reg. No. 40,629

Attorneys for Petitioners Amazon.com, Inc.,
Amazon.com LLC, Amazon Web Services,
Inc., A2Z Development Center, Inc. d/b/a
Lab126, Rawles LLC, AMZN Mobile LLC,
AMZN Mobile 2 LLC, Amazon.com Services,
Inc. f/k/a Amazon Fulfillment Services, Inc.,
and Amazon.com Services LLC (formerly
Amazon Digital Services LLC)

CERTIFICATE OF SERVICE (37 C.F.R. §§42.63(e), 42.105(a))

The undersigned hereby certifies that a complete and entire copy of the above-captioned “**PETITION FOR *INTER PARTES* REVIEW UNDER 35 U.S.C. §§ 311-319**” and all supporting exhibits numbered 1001 – 1003, 1008 – 1104, 1025 – 1028, and 1030 were provided via Federal Express, postage prepaid, to the Patent Owner’s correspondence address as stated below:

Sheppard Mullin Richter & Hampton LLP
650 Town Center Drive, 10th Floor
Costa Mesa, CA 92626

*Patent Owner’s correspondence
address of record for USP 8,073,681*

FENWICK & WEST LLP

/J. David Hadden/

J. David Hadden

Reg. No. 40,629

Attorneys for Petitioners Amazon.com, Inc.,
Amazon.com LLC, Amazon Web Services,
Inc., A2Z Development Center, Inc. d/b/a
Lab126, Rawles LLC, AMZN Mobile LLC,
AMZN Mobile 2 LLC, Amazon.com Services,
Inc. f/k/a Amazon Fulfillment Services, Inc.,
and Amazon.com Services LLC (formerly
Amazon Digital Services LLC)

Date: July 28, 2020
Fenwick & West LLP
801 California Street
Mountain View, CA 94041