

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM SERVICES LLC,

Petitioner,

v.

VB ASSETS, LLC,

Patent Owner.

Case: IPR2025-01240

U.S. Patent No. 11,080,758

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 11,080,758**

TABLE OF CONTENTS

	Page
CHART OF CLAIMS	ix
I. OVERVIEW OF THE '758 PATENT	1
A. Alleged Invention	1
B. Priority Date	2
II. IDENTIFICATION OF CHALLENGE	3
A. Statutory Grounds.....	3
B. Prior Art.....	4
1. Aretoulaki.....	4
2. Ramer	6
3. Kennewick.....	7
4. Hao	9
5. Jong	10
III. LEVEL OF ORDINARY SKILL IN THE ART	11
IV. CLAIM CONSTRUCTION	11
V. Ground 1: Aretoulaki and Ramer	11
A. Claim 1	11
[1.pre]	11
[1.a]	14
[1.b]	14
[1.c]	15
[1.d]	16
[1.e]	16

TABLE OF CONTENTS
(Continued)

	Page
[1.f]	17
[1.g]	20
[1.h]	23
B. Claim 2	26
[2.pre]	26
[2.a]	26
[2.b]	26
[2.c]	26
C. Claim 3	28
[3.pre]-[3.a]	28
[3.b]	29
[3.c]	29
D. Claim 4	30
E. Claim 5	30
[5.pre]	30
[5.a]	30
[5.b]	31
F. Claim 9	31
G. Claim 10	32
H. Claim 21	33
I. Claim 22	33

TABLE OF CONTENTS
(Continued)

	Page
J. Claim 23	33
K. Claims 24-27, 31-32, and 43-44	34
VI. Ground 2: Aretoulaki, Ramer, and Jong	34
A. Claim 8	34
B. Claim 30	35
VII. Ground 3: Aretoulaki, Ramer, and Kennewick	35
A. Claim 6	35
B. Claim 7	39
C. Claim 11	39
[11.pre]-[11.a]	39
[11.b]	40
D. Claim 12	40
E. Claim 13	40
F. Claim 14	41
G. Claim 15	41
H. Claim 16	41
I. Claim 17	42
[17.pre]-[17.a]	42
[17.b]	43
J. Claim 18	43
[18.pre]-[18.a]	43
[18.b]	44

TABLE OF CONTENTS **(Continued)**

	Page
[18.c]	44
[18.d]	44
K. Claim 19	46
[19.pre]-[19.a]	46
[19.b]	46
L. Claim 20	46
M. Claims 28-29 and 33-42	47
VIII. Ground 4: Kennewick and Hao	47
A. Claim 1	47
[1.pre]	47
[1.a]	50
[1.b]	51
[1.c]	51
[1.d]	52
[1.e]	55
[1.f]	56
[1.g]	58
[1.h]	61
B. Claim 2	65
[2.pre]	65
[2.a]	66

TABLE OF CONTENTS
(Continued)

	Page
[2.b]	66
[2.c]	67
C. Claim 4	68
D. Claim 5	68
[5.pre]	68
[5.a]	70
[5.b]	70
E. Claim 6	72
F. Claim 7	73
G. Claim 9	74
H. Claim 10	74
I. Claim 11	75
[11.pre]-[11.a]	75
[11.b]	75
J. Claim 12	75
K. Claim 13	76
L. Claim 14	77
M. Claim 15	77
N. Claim 16	77
O. Claim 17	77
[17.pre]-[17.a]	77
[17.b]	78

TABLE OF CONTENTS
(Continued)

	Page
P. Claim 18	78
[18.pre]-[18.a]	78
[18.b]	78
[18.c]	78
[18.d]	78
Q. Claim 19	80
[19.pre]-[19.a]	80
[19.b]	80
R. Claim 20	80
S. Claim 21	80
T. Claim 22	81
U. Claim 23	81
V. Claims 24, 26-29, 31-34	82
IX. Ground 5: Kennewick, Hao, and Aretoulaki	82
A. Claim 3	82
[3.pre]-[3.a]	82
[3.b]	84
[3.c]	85
B. Claim 25	86
X. Ground 6: Kennewick, Hao, And Jong	87
A. Claim 8	87
B. Claim 30	88

TABLE OF CONTENTS
(Continued)

	Page
XI. CONCLUSION.....	88
XII. STANDING.....	88
XIII. MANDATORY NOTICES AND FEES	88
A. Real Party-in-Interest	88
B. Related Matters.....	88
C. Lead and Backup Counsel.....	90
D. Service Information.....	90
E. Fees.....	91

TABLE OF AUTHORITIES

	Page(s)
STATUTES AND RULES	
35 U.S.C. § 102	4
35 U.S. C. § 103	3

EXHIBIT LIST (37 C.F.R. § 42.63(e))

Exhibit	Description
1001	U.S. Patent No. 11,080,758 (“the ’758 patent”)
1002	Declaration of Padhraic Smyth, Ph.D.
1003	Prosecution File History for the ’758 patent
1004	WO Pub. No. 02/073331 (“Aretoulaki”)
1005	U.S. Patent No. 7,752,209 (“Ramer”)
1006	U.S. Patent App. Pub. No. 2004/0193420 (“Kennewick”)
1007	U.S. Patent App. Pub. No. 2008/0189187 (“Hao”)
1008	U.S. Patent No. 6,173,250 (“Jong”)
1009	U.S. Patent No. 7,376,586 (“Partovi”)
1010	Asthana, A. & Krzyzanowski, P., “A Small Domain Communications System for Personalized Shopping Assistance,” Proceedings of 1994 International Conference on Personal Wireless Communications (pp. 199-203), IEEE Press (1994, August), (“Asthana”)
1011	Lucente, M., “Conversational interfaces for e-commerce applications,” Communications of the ACM (Sept. 2000), 43(9), 59-61, (“Lucente”)
1012	Huang et al., “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development,” Prentice Hall (2001) (“Huang”)
1013	U.S. Patent No. 6,859,776 (“Cohen”)
1014	U.S. Patent No. 7,640,160 (“Di Cristo”)
1015	Complaint for Patent Infringement, <i>VB Assets, LLC, v. Amazon.com Services, LLC</i> , 1:24-cv-00839-MN (D. Del. July 18, 2024), Dkt. 1
1016	Claim Construction Memorandum Order, <i>VB Assets, LLC, v. Amazon.com, Inc., et al.</i> , 1:19-cv-01410-MN (D. Del. June 23, 2021), Dkt. 90
1017	Curriculum Vitae of Padhraic Smyth, Ph.D.

CHART OF CLAIMS
[1.pre] A method for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon, the method being implemented by one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method, the method comprising:
[1.a] providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;
[1.b] receiving, by the one or more processors, words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine;
[1.c] determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;
[1.d] selecting, by the one or more processors, a purchase opportunity based on the determined context;
[1.e] delivering, by the one or more processors, the selected purchase opportunity via an electronic device in communication with the one or more processors;
[1.f] tracking, by the one or more processors, an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity, wherein the subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity;
[1.g] building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and
[1.h] interpreting, by the one or more processors, a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.
[2.pre] The method of claim 1, wherein the purchase opportunity relates to an item, the method further comprising:
[2.a] receiving, by the one or more processors, a second natural language utterance;

CHART OF CLAIMS
[2.b] determining, by the one or more processors, that the second natural language utterance comprises a request to purchase the item; and
[2.c] completing, by the one or more processors, a purchase transaction of the item based on the second natural language utterance.
[3.pre] The method of claim 1, the method further comprising:
[3.a] receiving, by the one or more processors, an indication of an occurrence of a user interaction with the purchase opportunity, wherein the interaction pattern includes the user interaction;
[3.b] determining, by the one or more processors, whether the interpretation of the natural language utterance was correct based on the indication; and
[3.c] interpreting, by the one or more processors, a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct.
[4.pre] The method of claim 1, wherein delivering the selected purchase opportunity via the electronic device comprises:
[4.a] audibly presenting, by the electronic device, a natural language conversational purchase opportunity.
[5.pre] The method of claim 1, wherein the computer program instructions comprise at least a conversational language processor configured to interpret a natural language utterance, the method further comprising:
[5.a] providing, by the one or more processors, the words or phrases as an input to the conversational language processor; and
[5.b] receiving, by the one or more processors, an interpretation of the natural language utterance, based on the words or phrases, as an output of the conversational language processor, wherein the purchase opportunity is selected based further on the interpretation of the natural language utterance.
[6.pre] The method of claim 5, wherein the conversational language processor comprises one or more domain agents, where a domain agent is configured to assist in:

CHART OF CLAIMS
[6.a] (i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.
[7] The method of claim 6, wherein the domains in which the one or more domain agents are configured to assist comprises an electronic commerce domain.
[8.pre] The method of claim 1, wherein recognizing the words or phrases from the natural language utterance comprises:
[8.a] using, by the speech recognition engine, phonetic dictation to recognize a stream of phonemes in the natural language utterance, wherein the recognized words or phrases are based on the recognized stream of phonemes.
[9.pre] The method of claim 1, the method further comprising:
[9.a] using, by the one or more processors, an environmental model to determine environmental information, wherein the context for the natural language utterance is determined based further on the environmental information.
[10] The method of claim 1, wherein the environmental information comprises a user location, a user activity, or a user action.
[11.pre] The method of claim 1, further comprising:
[11.a] invoking, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;
[11.b] wherein the one or more system resources include at least an agent comprising an autonomous executable, wherein the executable comprises a redistributable package of functionality for a specific application that receives, processes, and responds to user questions, queries, and/or commands.
[12] The method of claim 11, wherein the agent returns results of user questions, queries, and/or commands as responses to the user; and the responses are created using the results of information queries, system personality, and user preferences or other data in the user-specific profile.
[13] The method of claim 11, further comprising updating the agent from a remote location over a network to provide new functionality.

CHART OF CLAIMS	
[14]	The method of claim 11, wherein the agent uses system resources and the services of other agents.
[15]	The method of claim 11, further comprising providing a framework allowing the user to create or modify the agent.
[16]	The method of claim 14, further comprising providing a license management framework to enable the user to purchase the agent on a one time or subscription basis.
[17.pre]	The method of claim 11, wherein:
[17.a]	the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories;
[17.b]	the user can interact with, and receive the same services and applications from, any one of the devices on the network; and
[17.c]	the location of a particular device addressed by the user is used as part of context for questions asked of the user.
[18.pre]	The method of claim 1, further comprising:
[18.a]	capturing, by the one or more processors, the first natural language utterance;
[18.b]	parsing, by the one or more processors, the first natural language utterance;
[18.c]	interpreting, by the one or more processors, the first natural language utterance; and
[18.d]	invoking, by the one or more processors, an applicable agent from a library of agents.
[19.pre]	The method of claim 1, further comprising:
[19.a]	invoking an agent to process the natural language utterance; and
[19.b]	creating a query to a local database or through a network interface to a remote database.

CHART OF CLAIMS
[20] The method of claim 1, further comprising providing a user interface allowing the user to modify the user-specific profile, questions, queries, data sources, preferences, or parameters.
[21] The method of claim 1, wherein the context comprises a categorization of the recognized words or phrases.
[22] The method of claim 1, wherein the context comprises one of commands, directories, e-commerce, education, entertainment, lists, locations, multimedia, navigation, news, points of interest, search, stocks, recipes, scheduling, sports, subscriptions, transportation, or weather.
[23.pre] A system for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon, the system comprising:
[23.a] one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to:
[23.a (i)] provide a natural language utterance as an input to a speech recognition engine;
[23.a (ii)] receive words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine;
[23.a (iii)] determine a context for the natural language utterance based on the recognized words or phrases;
[23.a (iv)] select a purchase opportunity based on the determined context;
[23.a (v)] deliver the selected purchase opportunity via an electronic device in communication with the one or more processors;
[23.a (vi)] track an interaction pattern associated with the purchase opportunity delivered to the electronic device, wherein the tracked interaction pattern associated with the purchase opportunity includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity, wherein the subsequent request in which the electronic device interacts with the purchase opportunity includes the electronic device completing a transaction related to the purchase opportunity;

CHART OF CLAIMS
[23.a (vii)] build or update a user-specific profile based on the tracked interaction pattern; and
[23.a (viii)] interpret a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.
[24.pre] The system of claim 23, wherein the purchase opportunity relates to an item, and wherein the one or more processors are further programmed to:
[24.a] receive a second natural language utterance;
[24.b] determine that the second natural language utterance comprises a request to purchase the item; and
[24.c] complete a purchase transaction of the item based on the second natural language utterance.
[25.pre] The system of claim 23, wherein the one or more processors are further programmed to:
[25.a] receive an indication of an occurrence of a user interaction with the purchase opportunity, wherein the interaction pattern includes the user interaction;
[25.b] determine whether the interpretation of the natural language utterance was correct based on the indication; and
[25.c] interpret a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct.
[26.pre] The system of claim 23, wherein to deliver the selected purchase opportunity, the electronic device is configured to:
[26.a] audibly present a natural language conversational purchase opportunity.
[27.pre] The system of claim 23, wherein the computer program instructions comprise at least a conversational language processor configured to interpret a natural language utterance, wherein the one or more processors are further programmed to:

CHART OF CLAIMS
[27.a] provide the words or phrases as an input to the conversational language processor; and
[27.b] receive an interpretation of the natural language utterance, based on the words or phrases, as an output of the conversational language processor, wherein the purchase opportunity is selected based further on the interpretation of the natural language utterance.
[28.pre] The system of claim 27, wherein the conversational language processor comprises one or more domain agents, where a domain agent is configured to assist in:
[28.a] (i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.
[29] The system of claim 28, wherein the domains in which the one or more domain agents are configured to assist comprises an electronic commerce domain.
[30.pre] The system of claim 23, wherein to recognize the words or phrases from the natural language utterance, the speech recognition engine is configured to:
[30.a] use phonetic dictation to recognize a stream of phonemes in the natural language utterance, wherein the recognized words or phrases are based on the recognized stream of phonemes.
[31.pre] The system of claim 23, wherein the one or more processors are further programmed to:
[31.a] use an environmental model to determine environmental information, wherein the context for the natural language utterance is determined based further on the environmental information.
[32] The system of claim 23, wherein the environmental information comprises a user location, a user activity, or a user action.
[33.pre] The system of claim 23, wherein the one or more processors are further programmed to:
[33.a] invoke, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;

CHART OF CLAIMS
[33.b] wherein the one or more system resources include at least an agent comprising an autonomous executable, wherein the executable comprises a redistributable package of functionality for a specific application that receives, processes, and responds to user questions, queries, and/or commands.
[34] The system of claim 33, wherein the agent returns results of user questions, queries, and/or commands as responses to the user; and the responses are created using the results of information queries, system personality, and user preferences or other data in the user-specific profile.
[35] The system of claim 33, wherein the one or more processors are further programmed to update the agent from a remote location over a network to provide new functionality.
[36] The system of claim 33, wherein the agent uses system resources and the services of other agents.
[37] The system of claim 33, wherein the one or more processors are further programmed to provide a framework allowing the user to create or modify the agent.
[38] The system of claim 36, wherein the one or more processors are further programmed to provide a license management framework to enable the user to purchase the agent on a one time or subscription basis.
[39.pre] The system of claim 33, wherein:
[39.a] the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories;
[39.b] the user can interact with, and receive the same services and applications from, any one of the devices on the network; and
[39.c] the location of a particular device addressed by the user is used as part of context for questions asked of the user.
[40.pre] The system of claim 23, wherein the one or more processors are further programmed to:
[40.a] capture, by the one or more processors, the first natural language utterance;

CHART OF CLAIMS	
[40.b]	parse, by the one or more processors, the first natural language utterance;
[40.c]	interpret, by the one or more processors, the first natural language utterance; and
[40.d]	invoke, by the one or more processors, an applicable agent from a library of agents.
[41.pre]	The system of claim 23, wherein the one or more processors are further programmed to:
[41.a]	invoke an agent to process the natural language utterance; and
[41.b]	create a query to a local database or through a network interface to a remote database.
[42]	The system of claim 23, wherein the one or more processors are further programmed to provide a user interface allowing the user to modify the user-specific profile, questions, queries, data sources, preferences, or parameters.
[43]	The system of claim 23, wherein the context comprises a categorization of the recognized words or phrases.
[44]	The system of claim 23, wherein the context comprises one of commands, directories, e-commerce, education, entertainment, lists, locations, multimedia, navigation, news, points of interest, search, stocks, recipes, scheduling, sports, subscriptions, transportation, or weather.

U.S. Patent No. 11,080,758 (the “’758 patent”) claims technology for purchasing products using voice, but this technology was known in the prior art. *See, e.g.*, EX1001, 12:61-65; EX1002, ¶¶31-69. This Petition demonstrates that the prior art renders each ’758 patent claim obvious. *See also* EX1002, ¶¶1-344.

I. OVERVIEW OF THE ’758 PATENT

A. Alleged Invention

The ’758 patent is directed to “provid[ing] natural language processing based on advertisements” or “purchase opportunities.” EX1001, Abstract, 2:65-3:1; EX1002, ¶¶31-35. The ’758 patent identifies, as shown in FIG. 1, a system 100 that includes an input device 105, an Automatic Speech Recognizer 110 (“ASR”), a conversational language processor 120, and an output device 180, among others. EX1001, FIG. 1; EX1002, ¶33.

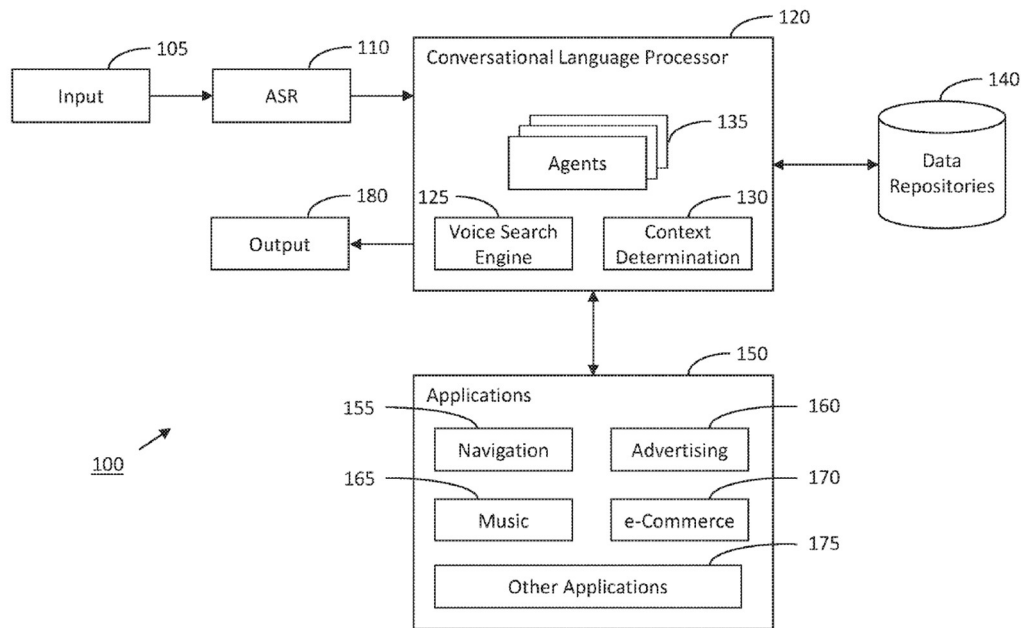


Figure 1

The system receives a user utterance at an input device and processes it by ASR 110. EX1001, 4:7-15; EX1002, ¶34. The conversational language processor “establish[es] a context” of the utterance and, based on a “context-based interpretation,” selects and presents a purchase opportunity to the user. EX1001, 5:4-9, 11:19-40; EX1002, ¶¶34-35. The system may “track historical data about users” to build user profiles and select purchase opportunities. EX1001, 11:12-18, 12:2-13; EX1002, ¶35.

B. Priority Date

The '758 patent was filed as Application No. 16/194,944 on November 19, 2018, and claims priority through a chain of continuations to Application

No. 11,671,526, filed on February 6, 2007. EX1001; EX1002, ¶31. This Petition treats February 6, 2007 as the priority date for the '758 patent.¹

II. IDENTIFICATION OF CHALLENGE

A. Statutory Grounds

Petitioner requests *inter partes* review and cancellation of claims 1-44 (the “challenged claims”) on the following six § 103 grounds. *See also* EX1002, ¶¶1-344.

Ground	Claims	Prior Art
1	1-5, 9-10, 21-27, 31-32, 43-44	Aretoulaki, Ramer
2	8, 30	Aretoulaki, Ramer, Jong
3	6-7, 11-20, 28-29, 33-42	Aretoulaki, Ramer, Kennewick
4	1, 2, 4-7, 9-24, 26-29, 31-44	Kennewick, Hao
5	3, 25	Kennewick, Hao, Aretoulaki
6	8, 30	Kennewick, Hao, Jong

¹ Petitioner reserves the right to challenge the '758 patent's priority date in this or other proceedings.

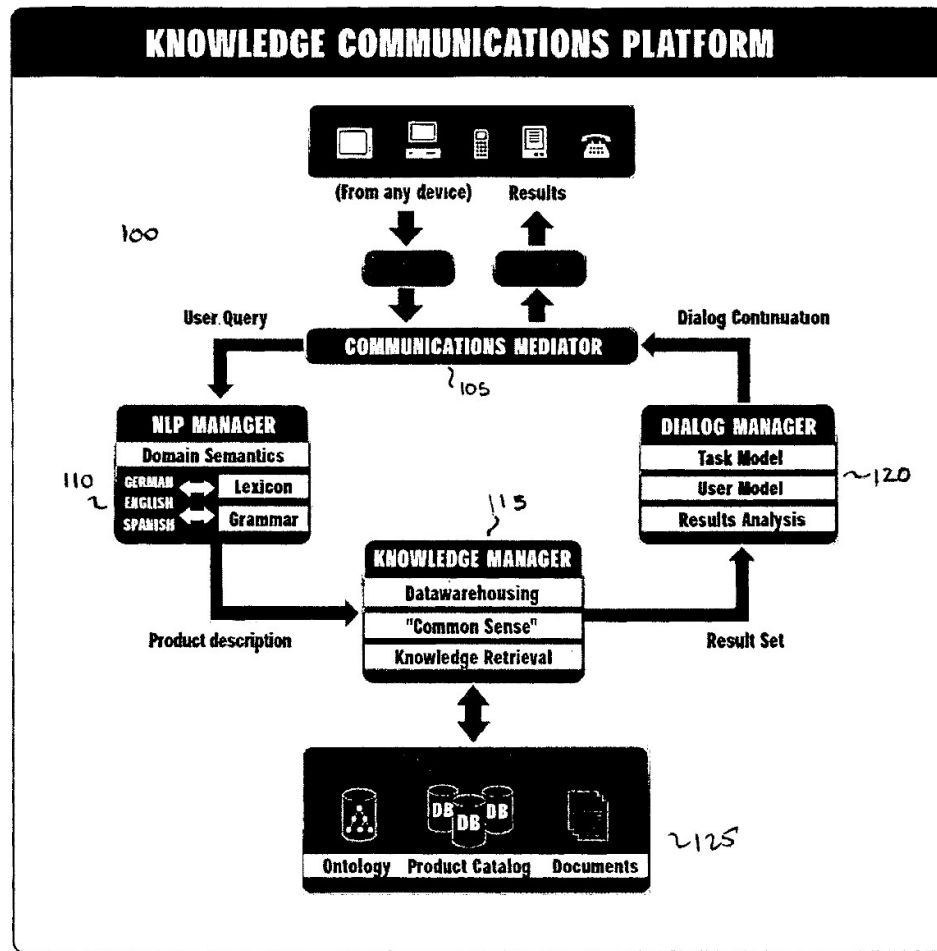
B. Prior Art²

1. Aretoulaki

Aretoulaki is prior art because it was filed on February 20, 2002 and published on September 19, 2002, before the '758 patent's priority date. EX1004; EX1002, ¶¶51.

Aretoulaki discloses a “knowledge communications platform” that enables “natural language (text or speech) dialogue-based electronic commercial transactions and information exchanges.” EX1004, Abstract; EX1002, ¶52. The platform, illustrated in Figure 1, includes the following “processing modules”: “Communications Mediator, 105,” “NLP Manager, 110,” “Knowledge Manager, 115,” and “Dialogue Manager, 120.” EX1004, 7:50-8:8, FIG. 1; EX1002, ¶52.

² Pre-AIA 35 U.S.C. § 102 applies to the '758 Patent.



Areoulaki's "Communications Mediator" receives user speech input 205 and translates it into a "Text Representation 215" routed "to NLP Manager 110 for further processing." EX1004, 8:9-18, FIG. 2; EX1002, ¶53. The NLP Manager analyzes the text representation to determine a domain associated with the received utterance to acquire a "domain semantic representation 225" identifying "the desired product, service or information." EX1004, 11:31-42; EX1002, ¶53. The Dialogue Manager and Knowledge Manager may perform further "discourse" and "contextual analysis" to validate and disambiguate the "domain semantic representation" using

prior discourse history and domain knowledge. EX1004, 12:15-52; EX1002, ¶53. For example, if there are “ambiguities,” the system generates “different interpretations...representing the various alternative meanings” associated with different domains, and uses discourse history, user models, and/or “salient domain concepts in the knowledge base” to weigh and select an interpretation. EX1004, 15:1-29; EX1002, ¶53. The Knowledge Manager uses the resulting “product (or service) description” to retrieve a “Result Set” of products/services from a database to present to the user (260-290). EX1004, 8:27-46, 9:12-20, 13:37-40, 15:44-51; EX1002, ¶54. The user can then “ask to make a purchase” of the product/service. EX1004, 22:20-23; EX1002, ¶54.

2. *Ramer*

Ramer is prior art because it was filed on January 19, 2006, before the '758 patent's priority date. EX1005; EX1002, ¶55.

Ramer discloses a “wireless search platform” in which users submit queries through “a voice-based query entry,” and the system returns results by selecting and presenting products available for purchase. EX1005, 15:47-49, 53:3-6, 55:4-9; EX1002, ¶¶56-57. The user can then select a product to purchase providing “spoken commands,” like “buy it now.” EX1005, 21:25-28 (“voice recognition facility” enables the system to “carry out spoken commands”); 47:63-48:13 (search results include “link to a commercial item” that enables “instant purchasing (e.g., ‘buy it

now’)); 42:21-24 (equating “voice-initiated actions” as “user interface equivalents” of “clicks and clickthroughs”); EX1002, ¶57.

Ramer’s system includes a “disambiguation facility” configured to interpret “ambiguous user requests” using information from a “mobile subscriber characteristics database” storing “user preferences” derived “through interpretation of other transactions...related to the mobile communication facility.” EX1005, 6:21-27, 17:28-48, 17:61-18:4, 23:60-24:16; EX1002, ¶58. The “queries may be disambiguated, or results returned, sorted, presented, or routed, based on [the user’s] shopping habits.” EX1005, 31:37-44, 34:4-21; EX1002, ¶58. For example, responsive to a query for “King,” “a frequent book purchaser may be presented with Stephen King books..., while a frequent music purchaser may be presented with results related to Elvis Presley.” EX1005, 34:4-21.

3. *Kennewick*

Kennewick is prior art because it was filed on July 15, 2003 and published on September 30, 2004, before the ’758 patent’s priority date. EX1006; EX1002, ¶59.

Kennewick discloses a “speech interface” enabling “mobile users to ask natural language speech questions or give natural language commands in a wide range of domains.” EX1006, [0002]; EX1002, ¶60. The speech interface makes “significant use of context...and user specific profile data” to interpret a user’s request and provide responses. EX1006, Abstract; EX1002, ¶60. As shown in

Figure 5, a user's natural language utterance is processed by a speech recognition engine 120. EX1006, FIG. 5, [0155]; EX1002, ¶60.

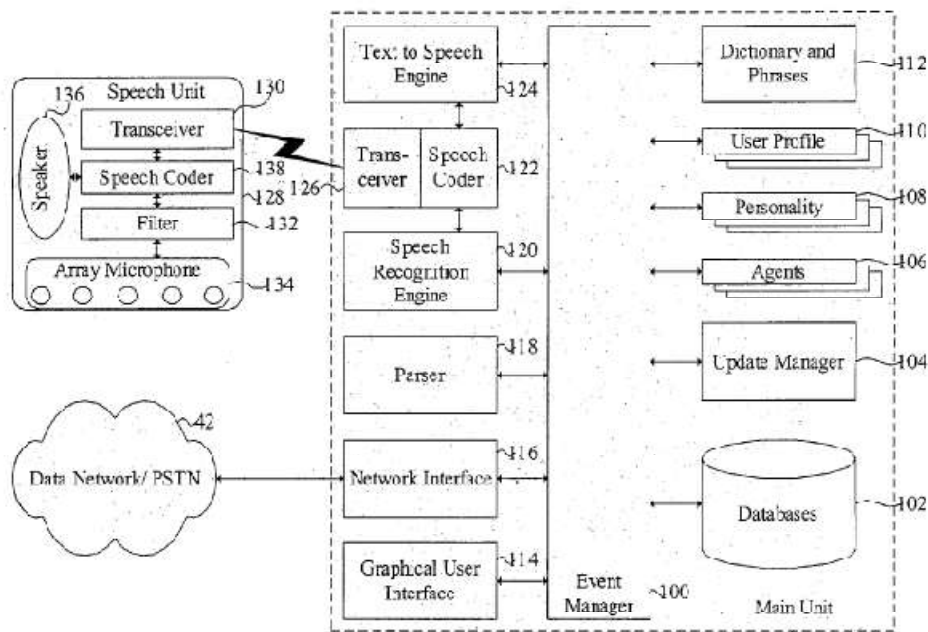


Figure 5. Speech Processing System Block Diagram

The system further includes a parser 118 that determines a context for the user's utterance based on the recognized words and phrases. EX1006, [0160]; EX1002, ¶61. The parser may “use a scoring system to determine the most[] likely context or domain [by]...weighting a number of factors including, the user profile 110, the domain agent's 156 knowledge[,] and previous context.” EX1006, [0160]-[0161]; EX1002, ¶61.

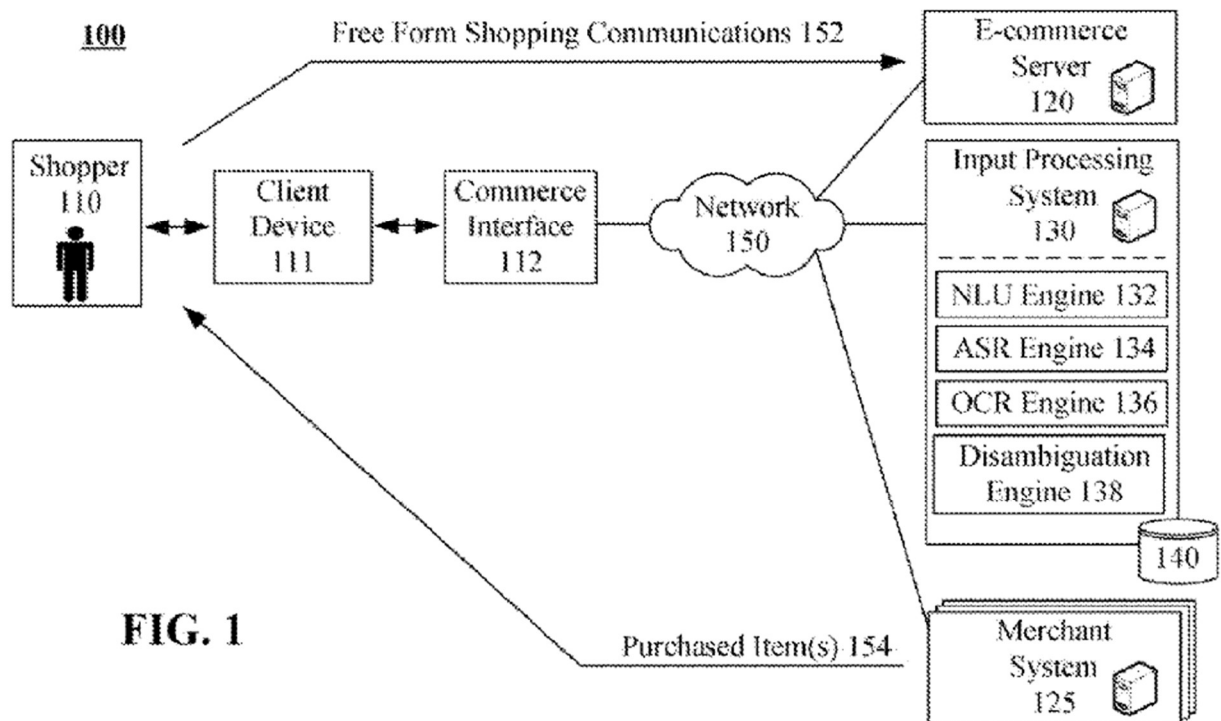
Kennewick's system may be used for voice commerce applications providing “[l]ocation-based marketing programs” and/or “[r]emote ordering and payment,” where the system “provide[s] interactive offers and promotions for goods and

services” and facilitates “remote ordering by using location information, customer preferences, customer order history, etc.” EX1006, [0060]-[0085]; EX1002, ¶62.

4. *Hao*

Hao is prior art because it was filed on February 1, 2007, before the ’758 patent’s priority date. EX1007; EX1002, ¶63.

Hao discloses “e-commerce shopping method[s]” where “shoppers can interact with an automated system using natural language input.” EX1007, Abstract, [0009]; EX1002, ¶64. Hao’s FIG. 1 shows the system including “an automatic speech recognition (ASR) engine 134” and a “natural language understanding (NLU) engine 132” to process speech input. EX1007, FIG. 1, [0022]; EX1002, ¶64.



The NLU Engine extracts “semantic meanings” from the user input, and a commerce server compares the extracted semantic meanings “to [available] inventory items...from at least one merchant system 125,” and adds matching items “to an electronic shopping cart” where the shopper can “confirm a purchase via options provided by the commerce interface.” EX1007, [0022]; EX1002, ¶65.

Hao further discloses that the user’s natural language input can contain “ambiguities,” which the system resolves “using shopper specific information maintained within a shopper profile/preferences data store.” EX1007, [0009], [0033], [0041]; EX1002, ¶¶66-67. This automatic resolution of ambiguities reduces a need to “prompt[] the shopper 110 to clarify their intended meaning,” streamlining the e-commerce process. EX1007, [0028]; EX1002, ¶67.

5. *Jong*

Jong is prior art because it was filed on June 3, 1998 and issued on January 9, 2001, before the ’758 patent’s priority date. EX1008; EX1002, ¶68.

Jong discloses a speech recognition device that translates “speech input signals...into text based on phonemes obtained from...the speech input signals.” EX1008, Abstract; EX1002, ¶69. The device identifies syllables based on the phonemes, and then groups the syllables into recognized words. EX1008, 5:51-67; EX1002, ¶69.

III. LEVEL OF ORDINARY SKILL IN THE ART

A POSITA at the priority date of the '758 patent would have had a Bachelor's degree in computer science, computer engineering, electrical engineering, or a related field in computing technology, and two years of experience with automatic speech recognition and natural language understanding, or equivalent education, research experience, or knowledge. Additional experience can substitute for the level of education, and vice-versa. EX1002, ¶¶1-13, 22-26.

IV. CLAIM CONSTRUCTION

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

V. GROUND 1: ARETOULAKI AND RAMER

A. Claim 1

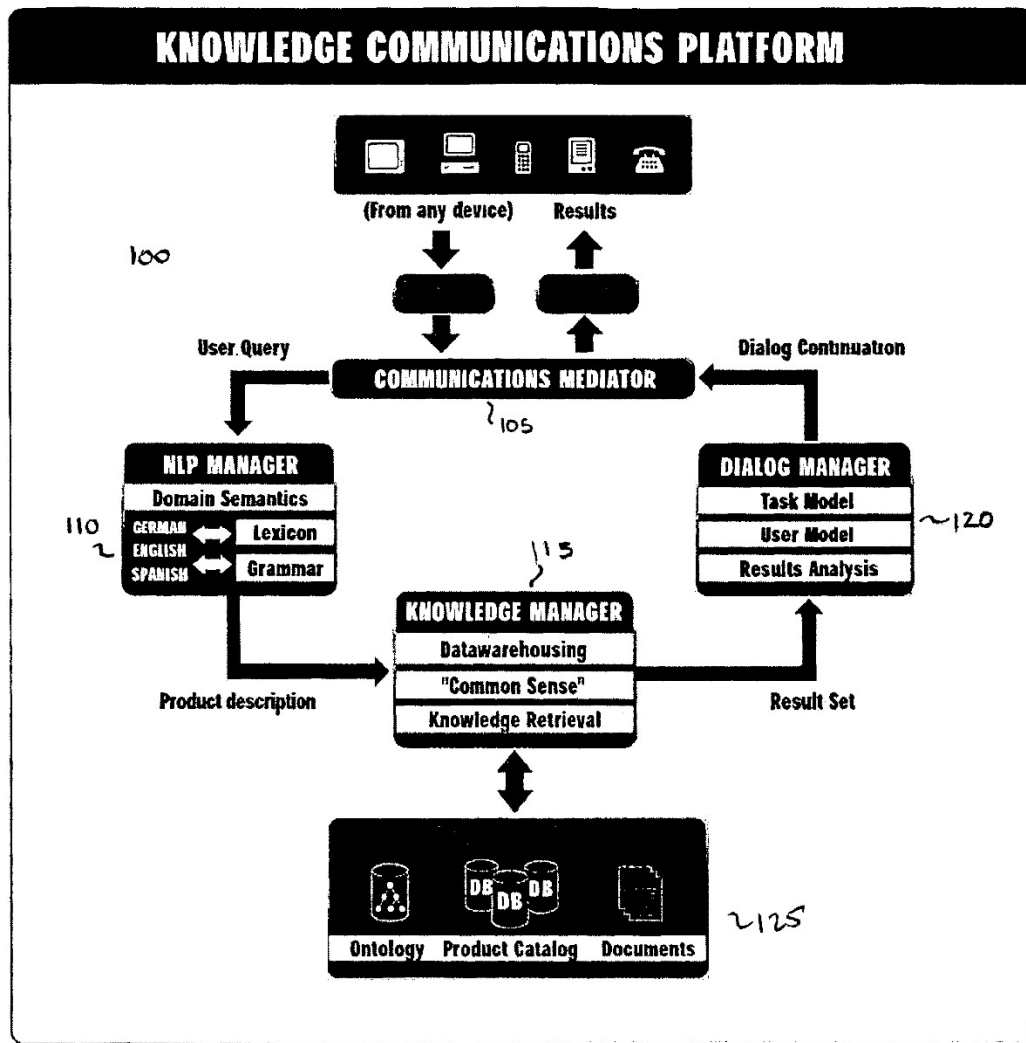
[1.pre]

Aretoulaki discloses a “knowledge communications platform” that performs a “*method for processing natural language utterances that include requests*” to conduct “natural language...dialogue-based electronic commercial transactions.” EX1004, Abstract, 8:27-46; EX1002, ¶¶76-77. Aretoulaki's system processes “natural language...speech input”/“user requests” to determine a “product (or

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

service) description” for selecting and presenting a “result set” of purchasable products/services (“*purchase opportunities*”) based thereon. EX1004, Abstract, 4:9-15, 8:27-46, 9:12-20, 15:44-51; EX1002, ¶¶76-77.

As shown in FIG. 1, Aretoulaki’s system includes “processing modules” like a “Communications Mediator,” “NLP Manager,” “Knowledge Manager,” and “Dialogue Manager,” which a POSITA would have found obvious to be implemented on “*one or more physical processors*” operating based on computer code, e.g., “*computer program instructions*,” which when executed cause the processing modules to perform the claimed steps. EX1004, 7:50-8:8; EX1002, ¶78.



In addition, a POSITA would have found it obvious for Aretoulaki in combination with Ramer to include “*one or more physical processors.*” EX1002, ¶79. Ramer discloses a “wireless search platform” for “processing [] user-initiated [voice] queries” deployed on “hardware and software,” e.g., a cluster of web servers, which a POSITA would have understood corresponds to “*one or more physical processors*” because servers contain physical processors. EX1005, 4:7-10, 15:47-49, 127:50-128:13; EX1002, ¶79. In addition, a POSITA would have understood that

software may be implemented as “*computer program instructions*” that when executed cause the processors to perform the claimed steps. In the combined system, a POSITA would have found it obvious to deploy Aretoulaki’s “processing modules” in web servers as taught by Ramer. EX1002, ¶79.

[1.a]

Aretoulaki’s “processing modules” corresponding to “*one or more processors*” include a “Communications Mediator” that receives a speech-based “natural language input” (e.g., a “*natural language utterance*”) and converts it “into text form.” EX1004, 6:17-21, 7:50-8:16, 45:24-27; EX1002, ¶80. A POSITA would have understood that the Communications Mediator includes the claimed “*speech recognition engine*,” to which it “*provid[es]*” the received natural language utterance for conversion “into text.” See EX1004, 5:31-37, 8:16-18, 10:36-39; EX1002, ¶80.

[1.b]

Aretoulaki’s “Communications Mediator” includes the “*speech recognition engine*” that generates an “ASCII text representation” corresponding to “*words or phrases, recognized from the natural language utterance*” and routes it “to NLP Manager 110 for further processing.” EX1004, 8:9-18, 10:36-39; EX1002, ¶81. The NLP Manager thus “*receiv[es]*” the text representation as output of the “*speech recognition engine*.” EX1004, 5:31-37, 8:9-18, 10:36-39; EX1002, ¶81.

[1.c]

Aretoulaki's NLP Manager processes the "text representation" to generate a "domain semantic representation" of the natural language utterance indicating "the [user's] desired product [or] service." EX1004, 11:31-42; EX1002, ¶82. To generate the "domain semantic representation," the NLP Manager determines a domain or application of the utterance (e.g., "***determining...a context for the natural language utterance***") by identifying "individual words" "point[ing] to domain specifications," and uses "a mapping lexicon" to "translate[] application-specific words into the corresponding relevant concepts in the domain ontology." EX1004, 11:41-45; EX1002, ¶82. This is consistent with the '758 patent's disclosure of determining an utterance's context "by having *domain* agents 135 competitively generate a context-based interpretation of the utterance," "scoring possible interpretations and selecting a highest scoring interpretation." EX1001, 5:4-9 (emphasis added⁴); EX1002, ¶83. Aretoulaki similarly discloses determining the context/domain by generating "different interpretations...representing the various alternative meanings in the case of ambiguities," e.g., if the utterance contains keywords "relevant to more than a single domain," and selecting one having "the highest weight." EX1004, 15:16-29; EX1002, ¶83; *see also* EX1004, 15:16-29

⁴ All annotations and emphases are added unless otherwise noted.

(domains include “booking” and “pricing”); EX1001, 9:36-56 (contexts include applications/domains such as “Music,” “Navigation,” or “Search”).

[1.d]

Areoulaki’s processing modules process the text representation of the utterance based on the determined context/domain “to acquire a [domain semantic] representation of the user’s requirements...about the desired product [or] service,” for searching “descriptions of existing products [or] services” in an application database. EX1004, 11:31-40; EX1002, ¶84. The Knowledge Manager “*select[s]*” the “best matches” from the search as a “result set,” corresponding to the “*select[ed]...purchase opportunity*,” for presentation to the user, where the user has the opportunity “make a purchase (commit)” of the presented products/services. EX1004, 15:44-51, 22:10-26; EX1002, ¶84.

[1.e]

Areoulaki’s Communications Mediator “coordinates the presentation of” (“*deliver[s]*”) the “retrieved product [or] service” (selected “*purchase opportunity*”) to the user as “a text message on the screen, a spoken prompt over the phone, images and graphics, or a combination of all.” EX1004, 8:20-25, 25:9-14; EX1002, ¶¶85-86. The user receives the selected purchase opportunity “*via an electronic device*” in communication with the Communications Mediator, such as “a Personal

Computer (PC),” “phone,” or “wireless device.” EX1004, 4:4-7, 8:12-14, Fig. 1; EX1002, ¶86.

[1.f]

Aretoulaki’s Dialogue Manager “*track[s]...an interaction pattern associated with the [delivered] purchase opportunity*” by monitoring “all the exchanges in a specific interaction session with the user,” including those associated with any delivered purchase opportunities. EX1004, 9:47-10:5, 14:1-4, 22:28-34; EX1002, ¶¶87-88. For example, the user, upon reviewing a delivered purchase opportunity, may, through the electronic device, make a “*subsequent request*” to “purchase (commit)” a selected product/service associated with the delivered opportunity, e.g., via a natural language request such as “Great! I’d like to book that” or “Sounds good. How can I pay?” EX1004, 8:12-14, 22:20-26, 40:22-35; EX1002, ¶¶88-89. The “*electronic device interacts with the purchase opportunity*” when the request is received by the system through the electronic device. See EX1004, 8:12-14; EX1002, ¶¶89-90. When the user makes this request, the system enables the user to, through the electronic device, “*complet[e] a transaction*,” e.g., by redirecting the user to a “corresponding e-commerce site” to complete the purchase. EX1004, Abstract, 22:20-26, 25:9-14; EX1002, ¶90. The system tracks the interaction by recording the completed purchase as part of the user’s “previous orders” in the “user model.” See EX1004, 9:40-49, 14:4-8; EX1002, ¶90.

To the extent that Aretoulaki does not explicitly disclose “*tracking...an interaction pattern*” that includes a subsequent request “*completing a transaction related to the purchase opportunity*,” the combination with Ramer renders it obvious. EX1002, ¶¶91-93. Ramer’s “wireless search platform” interacts with the user through a “mobile communication facility” (“*electronic device*”). EX1005, 3:64-67, 5:10-25, 5:53-55; EX1002, ¶91. The system processes the user’s “voice-based query entry” to return query results of purchasable products (a “*purchase opportunity*”), and tracks “*an interaction pattern associated with the purchase opportunity*” by tracking user activities “such as purchasing a product online” to draw inferences on the user’s preferences. EX1005, 10:41-50, 47:63-48:13, 53:3-6, 55:4-9; *see also id.*, 8:52-56, 16:11-19; EX1002, ¶¶91-92. For example, the purchase opportunity may include a link “to an item on an electronic commerce site” enabling “various e-commerce features, such as...single click purchasing,...instant purchasing (e.g., ‘buy it now’), tracking of clicks or transactions.” EX1005, 48:4-13; EX1002, ¶¶91-92. Ramer thus discloses tracking the “on-line interactions of” the “*electronic device*” with the delivered purchased opportunity, by “tracking clicks, clickthroughs, queries” and “purchases.” EX1005, 124:5-21; EX1002, ¶92. The tracked interaction pattern includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity to

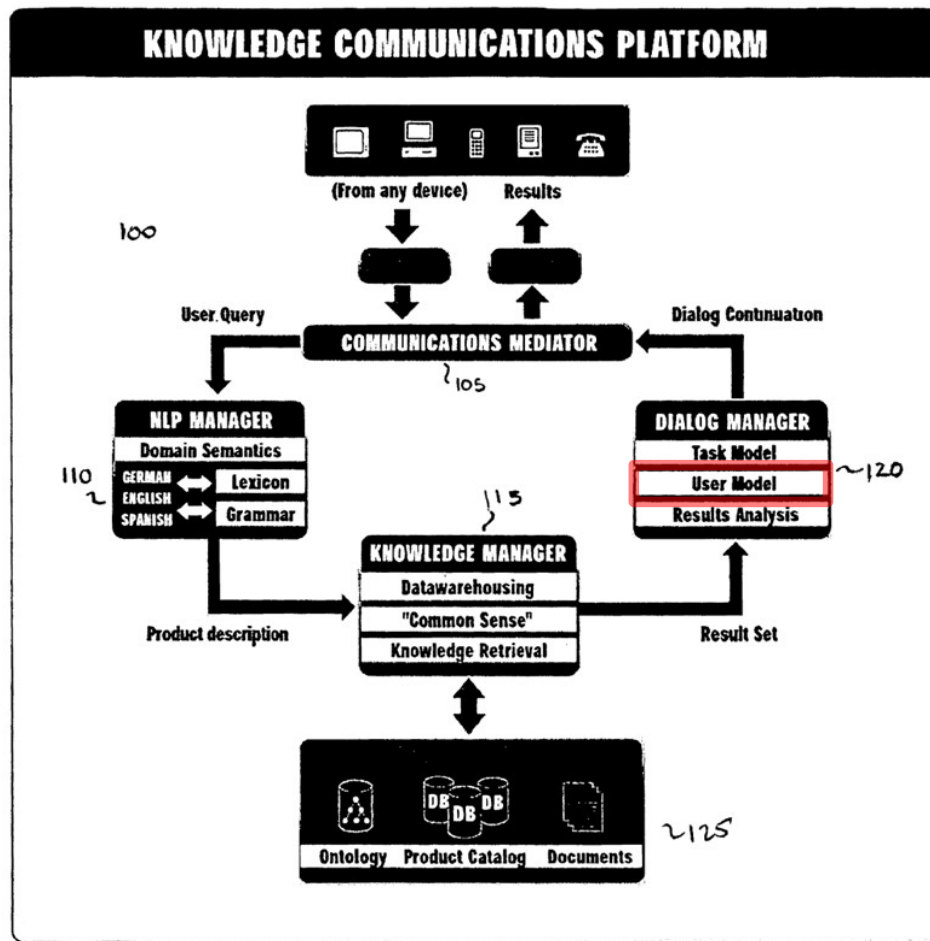
“complet[e] a transaction related to the purchase opportunity,” e.g., if the user purchases an associated item. EX1002, ¶92.

A POSITA would have been motivated to combine Aretoulaki and Ramer, as both references are in the same field of interactive speech technology and voice-based e-commerce through which a user can conduct “dialogue-based electronic commercial transactions.” EX1004, Abstract; *see also* EX1005, 4:7-10, 15:47-49; EX1002, ¶¶71-75, 93. A POSITA would have been motivated to modify Aretoulaki, to the extent not explicitly disclosed, to track user interactions including completing a purchase transaction as taught by Ramer. EX1002, ¶¶74, 93. For example, Aretoulaki uses the “user model” including the user’s “previous orders” and preferences to “interpret the user input” and “clarif[y]...ambiguities.” EX1004, 9:22-10:10, 14:4-18; EX1002, ¶¶74, 93. Ramer teaches tracking various interactions, including completed transactions (e.g., purchases), to register user preferences/“shopping habits,” to “deriv[e] greater clarity from ambiguous user queries.” EX1005, 12:21-24, 23:60-24:7, 34:4-21, 71:65-72:9; EX1002, ¶¶74, 93. A POSITA would have found this modification obvious as a combination of prior art elements according to known methods, to yield the predictable result of Aretoulaki’s system tracking interactions as taught in Ramer with which to “infer information and preferences about” the user. EX1004, 14:12-18; EX1002, ¶¶74-75, 93. A POSITA would have been motivated to make the modification to enhance and

improve Aretoulaki's ability to infer information about the user and "speed up the interaction and the search process." *Id.*

[1.g]

Aretoulaki's Dialogue Manager builds and updates a "user model contain[ing] data about the general preferences and assumptions of" the user based on monitored user behavior. EX1004, 9:39-10:10, 14:1-8, 22:28-34; EX1002, ¶94. The user model, shown in FIG. 1, is a "***user-specific profile***" built or updated "***based on the tracked interaction pattern***," as it includes a record of the user's "previous orders." EX1004, FIG. 1; EX1002, ¶¶94-95. A POSITA would have found obvious that the user model's record of the user's "previous orders" is updated whenever the user makes a new order by interacting with a presented purchase opportunity to purchase a product/service. EX1002, ¶¶94-95.



In addition, in the Aretoulaki-Ramer combination, Ramer discloses building or updating “mobile subscriber characteristics 112 pertaining to individual users” (“*user-specific profile*”) based on the user’s “transaction history” and “[u]ser preferences...derived from user behavior” including “clickthroughs, queries,” “products purchased,” and/or “other activities” (“*tracked interaction pattern[s]*”). EX1005, 2:45-63, 8:25-56, 10:34-50, 12:14-26, 124:13-21; EX1002, ¶¶96-97. For example, Ramer’s FIG. 3 illustrates the mobile subscriber characteristics 112 is built or updated using tracked data received from the wireless provider 108 and accessed

by the disambiguation facility 140 to disambiguate received queries. *See also* EX1005, 12:14-16, 23:60-24:27; EX1002, ¶¶96.

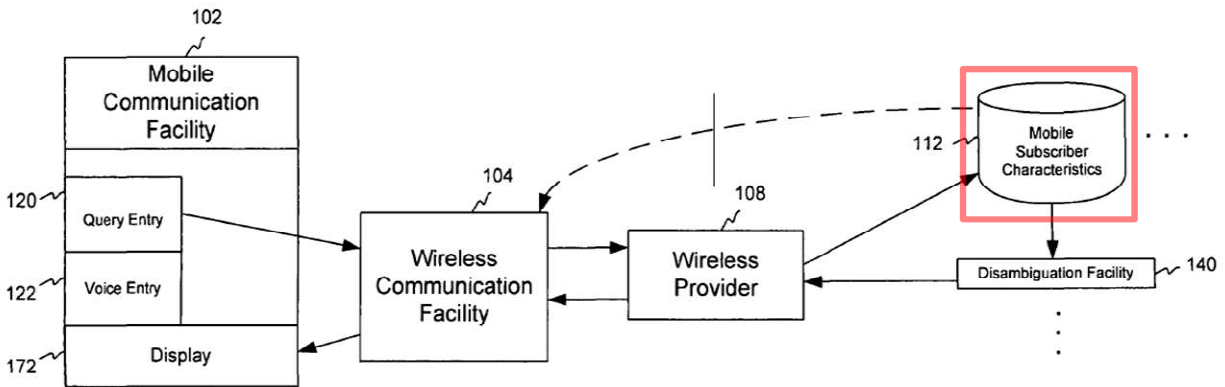


Fig. 3

A POSITA would have found it obvious to modify Aretoulaki, to the extent not explicitly disclosed, to update its “user model” based on tracked interaction patterns as taught by Ramer, by deriving user preference information “from [tracked] user behavior such as purchasing a product online.” *See* EX1005, 10:34-50; EX1004, 9:47-10:2; EX1002, ¶¶74, 97. A POSITA would have been motivated to make the modification to improve the utility of Aretoulaki’s user model for disambiguating user queries and “mak[ing] predictions” and “assumptions” about the user. *See* EX1004, 9:47-49, 14:12-18; EX1005, 12:14-16; EX1002, ¶¶74-75, 97. A POSITA would have found this modification obvious as a combination of prior art elements (e.g., Aretoulaki’s user model and Ramer’s mobile subscriber characteristics updated based on transaction history/user behavior) according to

known methods to yield predictable results (e.g., Aretoulaki's user models updated in the same way), to create a more robust system that improves Aretoulaki's ability to infer "what users are looking for" and "avoid[] tedious questioning of the user." EX1004, 6:50-7:2, 14:12-18; EX1002, ¶¶74-75, 97.

[1.h]

Aretoulaki's NLP Manager and Dialogue Manager "*us[e] the user-specific profile*" ("user model") containing the user's history of "previous orders" to "help...interpret" received utterances, including a "*subsequent natural language utterance*," by "infer[ring] information and preferences about" the user with which to set defaults and/or clarify ambiguities. EX1004, 9:39-10:10, 14:4-18; EX1002, ¶¶98-100. Based on the interpretation, the Knowledge Manager selects a "*subsequent purchase opportunity*" by "access[ing] the application database...to retrieve the best matches to the user's requirements as specified up to that point." EX1004, 15:44-51; EX1002, ¶¶98-100.

To the extent Aretoulaki does not explicitly disclose interpreting the subsequent natural language utterance using the user-specific profile "*to select [the] subsequent purchase opportunity*" as claimed, the combination with Ramer renders it obvious. EX1002, ¶¶101-104. Ramer discloses using "information from a user's profile," e.g., "products purchased," to interpret and "deriv[e] greater clarity from ambiguous user queries," and "make predictions regarding other

information/products that might interest the user.” EX1005, 12:14-24, 23:60-66; EX1002, ¶¶101-102. For example, responsive to a query for “King,” a user whose profile shows they are “a frequent book purchaser may be presented with Stephen King books,” whereas a user who is “a frequent music purchaser may be presented with results related to Elvis Presley.” EX1005, 34:18-21; *see also id.*, 24:65-25:17 (interpreting user query for “Coltrane” as relating to “the saxophonist, John Coltrane” based on “recent purchases”), 31:39-44; EX1002, ¶¶101-102.

A POSITA would have found it obvious to modify Aretoulaki to use user-specific profile information to interpret received user queries (“*subsequent natural language utterance*”) and select a purchase opportunity (“*subsequent purchase opportunity*”) based thereon as taught in Ramer. EX1002, ¶¶103-104. Both references disclose using user profile information to “infer information and preferences” about the user and clarify “ambiguities.” EX1004, 12:43-52, 14:4-18; EX1005, 23:60-66; EX1002, ¶¶74-75, 103-104. Ramer further provides specific examples of interpreting utterances using user-specific profile information to infer the user’s desired meaning and select appropriate purchase opportunities. *See* EX1005, 23:60-66, 31:39-44, 34:18-21; EX1002, ¶¶74-75, 103-104. For example, in the combination, a POSITA would have found it obvious for Aretoulaki’s system to interpret a “*subsequent natural language utterance*” including the term “King” by consulting a user-specific profile/model containing user preferences and

shopping habits inferred from previously-tracked interactions as taught by Ramer to select an appropriate “*subsequent purchase opportunity*,” e.g., books by Stephen King if the user is a frequent book purchaser. EX1004, 14:4-18; EX1005, 34:18-21; EX1002, ¶103. The combination would thus enable the system to better disambiguate the user queries and select an appropriate purchase opportunity for the user more closely matching “what [they] are looking for to satisfy their specific needs,” based on their past purchase history. EX1004, 4:48-7:2; EX1002, ¶¶75, 103-104.

A POSITA would have been motivated to combine the references to enhance Aretoulaki’s ability to use user-specific profile information to automatically interpret ambiguous user inputs and “avoid[] tedious questioning of the user.” EX1004, 14:12-18; EX1002, ¶¶74-75, 103-104. A POSITA would have had a reasonable expectation of success, as the modification involves a combination of prior art elements (e.g., Aretoulaki’s use of user profile information to interpret user requests, and Ramer’s specific examples of disambiguating user requests using user profile information such as past purchase history) according to known methods to yield predictable results (e.g., Aretoulaki’s system using Ramer’s disambiguation techniques to interpret ambiguous requests based on purchase history). EX1002, ¶¶74-75, 104.

B. Claim 2

[2.pre]

Areoulaki-Ramer discloses the delivered purchase opportunity relating to “*an item*” of “goods or services for sale,” such as a printer, scanner, skiing equipment, etc. EX1005, 47:63-48:12; EX1004, 6:52-7:2, 11:1-10, 13:18-35; EX1002, ¶105.

[2.a]

Areoulaki’s “Communications Mediator” receives user speech inputs, including a “*second natural language utterance*,” when the user, after viewing a purchase opportunity, “ask[s] to make a purchase” by saying “Great! I’d like to book that” or “Sounds good. How can I pay?” EX1004, 6:18-21, 8:9-16, 22:20-23, 40:22-35; EX1002, ¶106.

[2.b]

Areoulaki’s system processes each received utterance to determine a “user dialogue act[]” (“Dact”) corresponding to “the user intention,” such as determining that the user “ask[ing] to make a purchase” (“*second natural language utterance*”) corresponds to a “Commit” Dact indicating the user’s “*request to purchase the item*.” *Id.*, 12:30-35, 22:20-23, 23:21-23, 40:22-24; EX1002, ¶107.

[2.c]

Areoulaki’s system, responsive to the “*second natural language utterance*” corresponding to a “Commit” Dact, completes the purchase by directing the user “to

the corresponding E-commerce site.” EX1004, 25:9-14; EX1002, ¶108. This purchase functionality is “smoothly integrated in [Aretoulaki’s] architecture,” enabling the system to conduct “dialogue-based electronic commercial transactions.” EX1004, Abstract, 22:20-26, EX1002, ¶108.

To the extent that Aretoulaki does not explicitly disclose “*completing...a purchase transaction of the item*” in response to a user’s request to “purchase (commit),” the combination with Ramer renders it obvious. EX1002, ¶¶109-111. Aretoulaki’s stated purpose is providing “dialogue-based *electronic commercial transactions*,” which necessarily require completing a purchase transaction. EX1004, Abstract; EX1002, ¶108. And Ramer expressly discloses that the user may, after viewing presented products/services, “initiate...an online order” to complete the purchase. EX1005, 16:15-19, 136:33-39; EX1002, ¶¶109-110. The user can select a product to purchase via a voice command (“*second natural language utterance*”), such as “buy it now.” EX1005, 21:25-28 (“voice recognition facility” enables the system to “carry out spoken commands”); 47:63-48:13 (search results include “link to a commercial item” that enables “instant purchasing (e.g., ‘buy it now’)”); 42:21-24 (equating “voice-initiated actions” as “user interface equivalents” of “clicks and clickthroughs”); EX1002, ¶¶57, 109. The system provides “mobile wallet/billing-on-behalf-of” services to complete the transaction, e.g., by “having the wireless provider 108 collect the user’s credit card information as part of the

billing cycle and adding goods transactions automatically to the wireless provider's bill to the user." EX1005, 8:52-56, 124:54-56; EX1002, ¶¶109-110. A POSITA would have been motivated to combine Ramer's system for completing a purchase transaction with Aretoulaki to enable Aretoulaki to achieve its stated purpose of enabling "dialogue-based *electronic commercial transactions*." See EX1004, Abstract; EX1005, 16:15-19, 21:25-28; EX1002, ¶111. A POSITA would have found the combination obvious as a simple substitution of one known element for another to obtain predictable results because it simply modifies Aretoulaki's system so that the user's "*second natural language utterance*" indicating a "Commit" Dact functions as a "spoken command[]" for "initiat[ing]...an online order," causing the system to purchase and deliver the item as taught by Ramer. EX1002, ¶111.

C. Claim 3

[3.pre]-[3.a]

Aretoulaki discloses monitoring received "*user interaction[s] with [a] purchase opportunity*" "to locate communication or understanding problems," by detecting "[b]reaks in the normal flow, such as...objections on the part of the user" indicating "erroneous recognition on the part of the system," and "trigger[ing] a repair mechanism." EX1004, 7:11-19, 7:37-41, 18:19-25; EX1002, ¶¶112-114. The objection is a "*occurrence of a user interaction with the purchase opportunity*," which Aretoulaki's Communications Mediator receives "*an indication of*" via the

user's electronic device. *See* EX1004, 8:9-13; EX1002, ¶¶113-114. The received indication is part of the tracked interaction pattern, because Aretoulaki discloses “constantly monitoring”/tracking the user's interactions, including with any purchase opportunities. EX1004, 7:10-24; EX1002, ¶114.

[3.b]

Aretoulaki's Dialogue Manager determines “*whether the interpretation of the natural language utterance was correct*” by determining if the “*user interaction*” associated with “*the indication*” corresponds to a “Correct” Dact indicating the user “intends to correct the instantiation for a parameter” “because the system has misrecognized or misinterpreted a previous user input,” e.g., the “*natural language utterance.*” EX1004, 12:30-35, 38:44-51; EX1002, ¶115.

[3.c]

If the user's interaction corresponds to a “Correct” Dact, Aretoulaki's Dialogue Manager triggers a “repair mechanism” to correct a misinterpreted parameter of the natural language utterance. EX1004, 18:19-35, 32:21-25, 39:1-5; EX1002, ¶116. Because the system interprets each utterance “in terms of the preceding exchanges,” Aretoulaki then, following a repair correcting an incorrectly interpreted parameter, interprets “*subsequent natural language utterance[s] based on the determination of whether the interpretation of the [prior] natural language*

utterance was correct,” by interpreting them in the context of the now-corrected parameter. EX1004, 4:17-23, 16:38-40, 22:28-31; EX1002, ¶117.

D. Claim 4

Aretoulaki delivers the selected purchase opportunity via the electronic device “*audibly*” as “speech,” e.g., “*natural language conversational*” speech, enabling the system to conduct a “natural language dialogue-based electronic commercial transaction.” EX1004, 7:50-52, 15:44-51; EX1002, ¶118.

E. Claim 5

[5.pre]

Aretoulaki’s processing modules include an “NLP Manager,” “Dialogue Manager,” and “Knowledge Manager.” See EX1004, 7:50-8:7, FIG. 1; EX1002, ¶119. A POSITA would have understood that these processing modules operate using computer code (“*computer program instructions*”), and serve as a “*conversational language processor*” that interprets a natural language utterance, as discussed below. EX1004, 8:4-8, 15:44-51; EX1002, ¶119.

[5.a]

Aretoulaki’s Communications Mediator translates the natural language utterance into “an ASCII string” of recognized “*words or phrases*,” and “*provid[es]*” it as input to the NLP Manager (of the “*conversational language processor*”) “for further processing.” EX1004, 8:9-18, 10:36-39; EX1002, ¶120.

[5.b]

Areoulaki's "NLP Manager," "Dialogue Manager," and "Knowledge Manager," which form the "*conversational language processor*," receive and process the "*words or phrases*" to generate an "*interpretation of the natural language utterance*" in the form of "frames" specifying a "product (or service) description." EX1004, 8:12-18, 8:40-46 ("NLP Manager" generates "frames"), 9:6-12 ("Dialogue Manager" and "Knowledge Manager" further process frames to locate and resolve any "inconsistencies, incompatibilities, and contradictions"); EX1002, ¶¶121-122. The Knowledge Manager "*receiv[es]*" these frames, and uses them to select the purchase opportunity by "access[ing] the application database...to retrieve the best matches to the user's requirements." EX1004, 15:44-51; EX1002, ¶122.

F. Claim 9

Areoulaki's system maintains a "discourse history" meeting the claimed "*environmental model*," from which it determines previous "user input[s]" and "user dialogue acts," which is "*environmental information*," to "help it interpret the user input." EX1004, 9:39-10:10, 15:23-25; EX1002, ¶123; *see also* EX1001, claim 10. The system determines the context for the natural language utterance "*based...on the environmental information*" by using the discourse history's user dialogue act

information to weigh and select between interpretations corresponding to different domains/contexts. EX1004, 15:16-29; EX1002, ¶124.

In addition, Ramer discloses maintaining “mobile subscriber characteristics” (“*environmental model*”) containing user location information (“*environmental information*”), and using it to select between potential interpretations of the user’s utterance. EX1005, 2:56-63, 8:57-58, 17:64-18:16, 24:16-27, 28:6-8; EX1002, ¶125. A POSITA would have been motivated to modify Aretoulaki to incorporate Ramer’s consideration of user location when weighing “different interpretations” to create a system that considers more relevant information when interpreting user speech, thereby, “speed[ing] up the interaction...[and] avoiding tedious questioning of the user.” EX1004, 14:12-18, 15:16-29, EX1005, 24:16-27; EX1002, ¶126. A POSITA would have found this modification obvious as a combination of prior art elements (considering discourse history *or* user location to interpret user requests) according to known methods to yield predictable results (considering *both* discourse history *and* user location when interpreting user requests). EX1002, ¶126.

G. Claim 10

Aretoulaki-Ramer discloses “*environmental information*” comprising previous user inputs and dialogue acts (“*a user activity, or a user action*”), or a “*user location*.” See EX1004, 10:5-10, 15:23-25; EX1005, 2:56-63, 8:57-58, 17:64-18:16; EX1002, ¶127.

H. Claim 21

Aretoulaki discloses determining the context by identifying “individual words...point[ing] to domain specifications” and “*categoriz[ing]...the recognized words*” into “corresponding relevant concepts in the domain ontology.” EX1004, 4:48, 11:41-45; EX1002, ¶128.

I. Claim 22

Aretoulaki discloses that the context/domain includes those for “*e-commerce*” (e.g., travel booking/pricing, electronics purchasing, etc.) or “*transportation*” (e.g., travel booking/pricing). EX1004, 15:20-22, 20:29-34; EX1002, ¶129.

J. Claim 23

Claim 23 is rendered obvious for similar reasons as claim 1. EX1002, ¶130. The claimed “*system*” is met by Aretoulaki’s “Knowledge Communications Platform,” which as discussed above, includes processing modules that “*process[] natural language utterances that include requests*” to “*select[] and present[] purchase opportunities based thereon.*” EX1004, Abstract, 7:50-8:8, 8:27-46, FIG. 1; EX1002, ¶130. A POSITA would have found obvious that the processing modules include “*one or more physical processors programmed with computer instructions*” which when executed cause the processing modules perform the claimed steps. EX1004, 7:50-8:8; EX1002, ¶130. For example, a POSITA would have found it obvious for the processing modules to be implemented on web servers

as taught by Ramer. *See* EX1005, 127:50-128:13; EX1002, ¶130. The remaining elements of claim 23 are substantively identical to those of claim 1. EX1002, ¶130.

K. Claims 24-27, 31-32, and 43-44

Claims 24-27, 31-32, and 43-44 are substantively identical to claims 2-5, 9-10, and 21-22, respectively, and are obvious for the same reasons. EX1002, ¶¶131-138.

VI. GROUND 2: ARETOULAKI, RAMER, AND JONG

A. Claim 8

To the extent that Aretoulaki-Ramer does not explicitly disclose recognizing the words or phrases of the natural language utterance using “*phonetic dictation*,” the combination with Jong renders it obvious. EX1002, ¶¶139-140. While Aretoulaki-Ramer leaves open the specific technique it uses to recognize speech, Jong discloses a technique performed by a “speech recognition device” that translates “speech input signals into textual representations” corresponding to a stream of phonemes, i.e., “*phonetic dictation*.” EX1008, Abstract, 9:34-46; EX1002, ¶140. The “*recognized words or phrases*” are “*based on*” “word syllables” extracted from “*the recognized stream of phonemes*.” EX1008, 5:50-67; EX1002, ¶141. A POSITA would have been motivated to apply Jong’s phonetic dictation technique as a straightforward way to implement speech recognition in the Aretoulaki-Ramer system. EX1002, ¶¶142-143. The combination uses a known technique (Jong’s phonetic dictation) to improve a similar system (Aretoulaki-

Ramer, which leaves open the speech recognition technique it uses) in the same way.

Id.

B. Claim 30

Claim 30 is substantively identical to claim 8 and is obvious for the same reasons. EX1002, ¶144.

VII. GROUND 3: ARETOULAKI, RAMER, AND KENNEWICK

A. Claim 6

Aretoulaki's "*conversational language processor*" uses domain-specific information to "*interpret[] requests*" and "*determin[e] a response to the requests related to [a] domain*" by interpreting the utterance using "domain-specific grammars and lexica" to establish "the salient domain concepts and their values appearing in the user input," and accessing application databases to retrieve a "Result Set" based thereon. EX1004, 4:48-51, 8:50-52, 9:13-20, 13:37-45, 15:44-51; EX1002, ¶¶151-152.

To the extent that Aretoulaki-Ramer does not explicitly disclose the "*conversational language processor*" comprising "*one or more domain agents*" assisting in the claimed functions, the combination with Kennewick renders it obvious. EX1002, ¶¶145-150, 153-156. Kennewick discloses a "*conversational language processor*" that includes a "parser" and "agents," which as shown in FIG. 6, include "*domain agents.*" EX1006, [0123]; EX1002, ¶153.

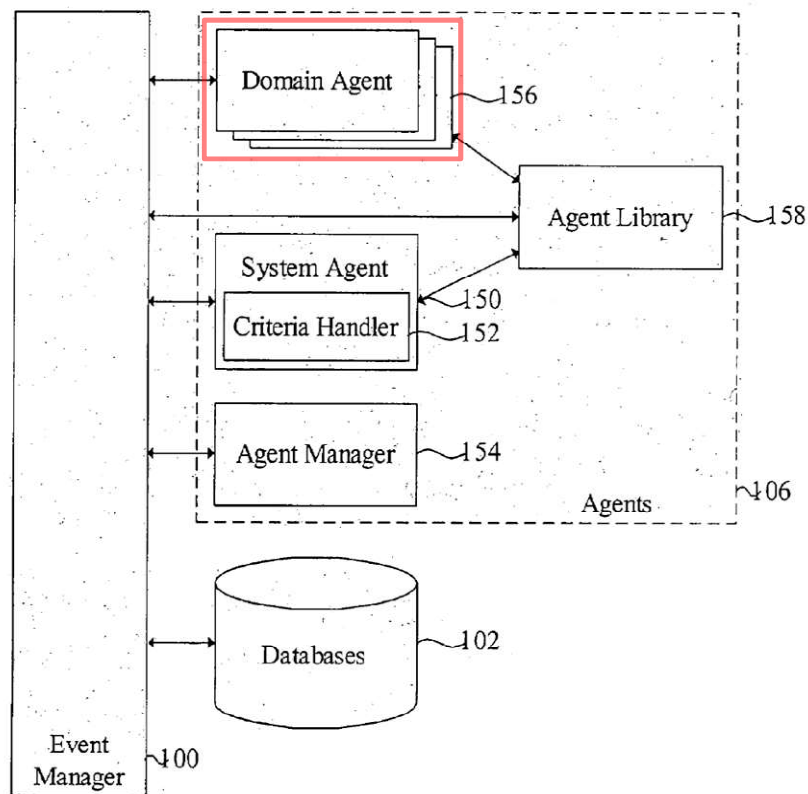


Figure 6. Agent Architecture

Each domain agent provides “behavior and data required for a specific area of functionality,” and assists the parser in “*interpreting requests related to its domain*” by “suppl[ying]” domain-specific information for the parser “to determine the most plausible interpretation” of an utterance. EX1006, [0018], [0029], [0123], [0163]; EX1002, ¶154. Each domain agent further assists in “*determining a response to the request relating to its domain*” by “create[ing] one or more queries to...information sources” related to its domain, for “create[ing] a satisfactory response to” the request. EX1006, [0167], [0178]; EX1002, ¶154.

A POSITA would have found it obvious to combine Aretoulaki-Ramer with Kennewick, because the references are in the same field of interactive speech technology and voice-based e-commerce. *See, e.g.*, EX1004, Abstract, 8:27-48, 22:10-23; EX1005, 4:7-10, 15:47-49, 47:34-39, 47:63-48:12; EX1006, [0012], [0014], [0066]-[0067]; EX1002, ¶¶145-150, 155-156. Additionally, Aretoulaki and Kennewick each teach determining a context/domain of a received utterance, and processing utterances using domain-specific information. *See, e.g.*, EX1004, 4:37-42, 8:50-52, 10:15-18; EX1006, [0011], [0017], [0032], [0160]-[0161]; EX1002, ¶¶147-148, 155-156. For example, Aretoulaki discloses generating a “domain semantic representation” of the user’s input, which the Knowledge Manager uses to search an appropriate “application database” for the “best matches to the user’s requirements.” EX1004, 8:50-52, 11:31-45, 15:46-51; EX1002, ¶¶147, 155. Similarly, Kennewick discloses determining a “most[] likely context or domain for a user’s question or command” to invoke an appropriate “domain agent” to “make one or more queries” based on the user request. EX1006, [0017], [0032], [0160]-[0161], [0167]; EX1002, ¶¶148, 155.

A POSITA would have been motivated to modify Aretoulaki’s Knowledge Manager, which “maintains and manipulates information on...specific domain[s] and application[s],” to utilize an “agent architecture” of individual domain agents to assist in interpreting requests and creating responses to requests relating to specific

domains as taught by Kennewick. EX1004, 8:50-52, 9:13-16; EX1006, [0126]; EX1002, ¶¶149-150, 155. Because each domain agent “include[s] everything needed to extend or modify the functionality of the system in a current or new domain,” the modification would create a more modular system more easily extended into new domains/applications. EX1006, [0126]; EX1002, ¶¶149, 155. Indeed, Aretoulaki already discloses that the Knowledge Manager accesses data of different applications and domains (e.g., application databases) through a “resource agent,” and it would have been obvious to organize the resource agent into different domain-specific agents managing resources for respective domains as taught by Kennewick, given Kennewick’s teaching that agents can use “the services of other, typically more specialized, agents.” See EX1004, 9:12-17; EX1006, [0017]; EX1002, ¶¶149, 155.

A POSITA would have had a reasonable expectation of success because the combination is a simple substitution of one known technique (managing domain-specific behavior and data for specific domains through a Knowledge Manager) for another (managing domain-specific behavior and data for specific domains through respective domain agents) to obtain predictable results (managing domain-specific behavior and data for specific domains through a Knowledge Manager using respective domain agents). EX1002, ¶¶150, 156. A POSITA would have found this modification obvious as a combination of prior art elements according to known

methods, and would have found the modification obvious to try to create a more modular and extensible system. *Id.*

B. Claim 7

Aretoulaki discloses enabling “electronic commercial transactions” via “*electronic commerce domain[s]*” for purchasing travel packages, electronics, etc., which the Knowledge Manager in the combination manages through respective domain agents as taught by Kennewick. EX1004, 7:50-52, 15:20-22, 20:29-34, 22:10-18; *see also* EX1006, [0018], [0067] (domains for “travel services, entertainment scheduling,” and “remote ordering and payment”); EX1002, ¶¶157-159.

C. Claim 11

[11.pre]-[11.a]

As discussed for claim 1, Aretoulaki’s processing modules “*select the purchase opportunity based on the determined context,*” where the Knowledge Manager uses the interpretation based on the determined context/domain to access an application database via a “resource agent” and select the purchase opportunity. EX1004, 9:12-17, 11:31-40, 15:44-51; EX1002, ¶¶160-161. The Knowledge Manager, resource agent, and/or application database are invoked “*system resources.*” EX1002, ¶161. In addition, in the combination discussed above where Aretoulaki’s Knowledge Manager is modified to utilize Kennewick’s agents,

Kennewick discloses “**invoking** the proper resources, including agents,” which are “**system resources**” available to the system, to “**select the purchase opportunity.**”

EX1006, [0017], [0028], [0162]; EX1002, ¶¶162-163.

[11.b]

Aretoulaki-Ramer-Kennewick invokes “**system resources includ[ing] at least an agent,**” e.g., a domain agent, which Kennewick discloses is “**an autonomous executable**” corresponding to a “**re-distributable package[]...of functionality...for a specific domain**” or “**application area,**” that “**receives, processes and responds to user questions, queries, and commands.**” EX1006, claim 1, [0017], [0126]; EX1002, ¶164.

D. Claim 12

In the combination discussed above where Aretoulaki’s Knowledge Manager invokes an agent based on the determined context/domain, Kennewick discloses that its “**agent**” “returns results of user questions” as “responses to users,” created “using the results of information queries, the system personality 108[,] and the user preferences or other data in the user[-specific] profile 110.” EX1006, [0124]; EX1002, ¶¶165-167.

E. Claim 13

Kennewick discloses an “update manager” for “**updating the agent**” “remotely over a network” to “add[] new behaviors,” which a POSITA would have

understood corresponds to “*new functionality*.” EX1006, [0017], [0098], [0120], [0126], [0129]; EX1002, ¶168. A POSITA would have found it obvious for the combination to update agents this way, to extend the functionality of the system. EX1002, ¶169.

F. Claim 14

In the combination, Kennewick discloses that the agent may use “*system resources and the services of other*, typically more specialized, *agents*.” EX1006, [0017]; EX1002, ¶170.

G. Claim 15

Kennewick discloses providing a “*framework*” such as speech or graphical interface allowing the user to “*create* new agents” and/or “*modify*...existing agents.” EX1006, [0134]; EX1002, ¶171. A POSITA would have found it obvious to incorporate this teaching in the combination, to allow users to “extend the system and configure it to their own preferences.” EX1006, [0021]; EX1002, ¶171.

H. Claim 16

In the combination, Kennewick discloses “*providing a license management framework*” by providing “license management capability allowing the sale of agents by third parties to one or more users *on a one time or subscription basis*.” EX1006, [0126], [0132]; EX1002, ¶172. A POSITA would have found it obvious to incorporate this teaching in the combination to enable users “to extend the

functionality of the invention to a new domain” by licensing new domain agents.
EX1006, [0017]; EX1002, ¶172.

I. Claim 17

[17.pre]-[17.a]

Aretoulaki’s system utilizes a “knowledge base” containing “expert domain knowledge” (“*data, information*”), user models (“*user-specific profiles*”), and discourse “*histories*.” See, e.g., EX1004, 4:17-42, 7:11-45, 8:1-8, 9:36-49, 10:5-10, 14:38-45; EX1002, ¶¶173-174. The user accesses the system over the Internet via any of a PC, phone, or PDA, which a POSITA would have understood to be “*a network of devices*” connected to a network. EX1004, Abstract, 3:51-4:7, 8:12-14, 10:31-33; see also EX1001, 4:7-11; EX1002, ¶¶174-175.

In addition, in the combination, Kennewick discloses that the system “can be deployed in *a network of devices using [a] common base of agents, data, information, user[-specific] profiles and histories*,” where the user can “*interact with, and receive the same services and applications* at any location equipped with the required device on the network,” e.g., from “*any one of the devices on the network*.” EX1006, [0016], claim 27; EX1002, ¶176. A POSITA would have found it obvious to modify Aretoulaki, to the extent not explicitly disclosed, to implement the method on “*a network of devices*” as taught by Kennewick, to increase the system accessibility to users via different devices at different locations. See also

EX1004, 10:31-33; EX1002, ¶177. A POSITA would have found this combination obvious, as it is a simple substitution of one known element (e.g., accessing the system through a client device) for another (e.g., accessing it through any of a network of devices) to obtain predictable results. EX1002, ¶177.

[17.b]

In the combination, Kennewick discloses that the “*network of devices*” are “placed at different locations,” and the system can “*use the location of the particular device addressed by the user as part of the context for the questions asked*” of the user. EX1006, [0016]; EX1002, ¶178. A POSITA would have found it obvious for the combination to use location as part of context as taught by Kennewick, to take advantage of a greater variety of information when interpreting received utterances. EX1002, ¶178. The modification is a combination of prior art elements (considering discourse history, and considering user location when interpreting user requests) according to known methods to yield predictable results (considering both when interpreting user requests). *Id.*

J. Claim 18

[18.pre]-[18.a]

Aretoulaki’s Communications Mediator “*captur[es]...the first natural language utterance*” by receiving speech input and converting it into text. EX1004,

8:9-18; *see also* EX1006, [0028] (“*capturing* the user's question or query through accurate speech recognition”); EX1002, ¶¶179-181.

[18.b]

Areoulaki’s “NLP Manager” functions as “the parser of the system” to “*pars[e]...the first natural language utterance*” by performing “lexical, syntactic (grammatical), and domain semantic analysis of the user input.” EX1004, 8:27-35, 28:44-48; *see also* EX1006, [0013] (“a parser for *parsing* the input”); EX1002, ¶¶182-183.

[18.c]

Areoulaki’s processing modules, as discussed for claim 5, “*interpret a natural language utterance,*” including the “*first natural language utterance.*” EX1004, 8:4-7; *see also* EX1006, [0028] (“parsing and *interpreting* the question or query”); EX1002, ¶¶184-185.

[18.d]

Kennewick discloses invoking “the required agent” to process the user’s request, selected from a library of agents 106. EX1006, [0028], [0032], [0127]-[0128]; EX1002, ¶186. For example, FIG. 6 illustrates Kennewick’s “agent architecture” having a collection or library of domain agents managed by an “agent manager 154.”

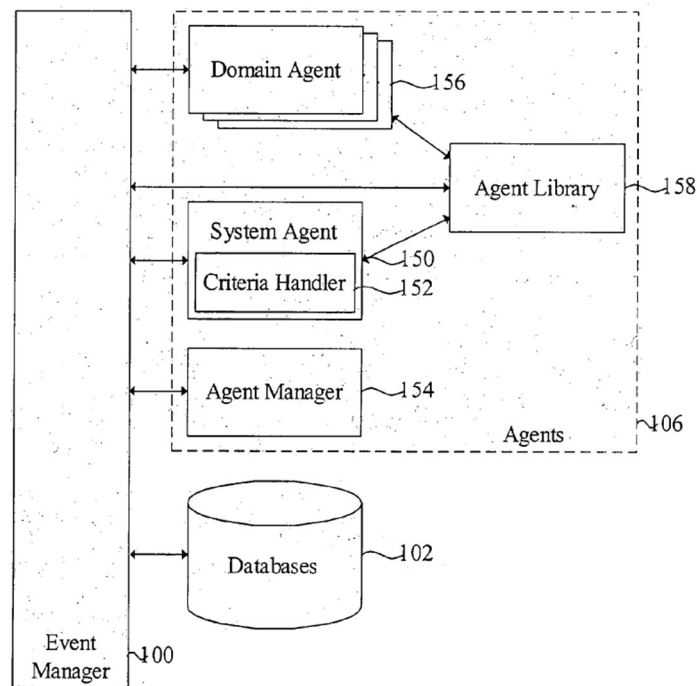


Figure 6. Agent Architecture

See EX1006, [0126], FIG. 6; EX1002, ¶¶186-187. A POSITA would have understood that Kennewick’s “required agent” is “*an applicable agent from a library of agents*,” because a POSITA would have found it obvious that a managed collection of agents to be a “library” of agents. EX1002, ¶¶186-187. In the combination where Aretoulaki’s Knowledge Manager is modified to utilize Kennewick’s agent architecture, it “*invok[es]...an applicable agent from*” the same “*library of agents*” corresponding to “the specific domain and application under consideration.” See EX1006, [0017], [0126]; EX1004, 8:50-52; EX1002, ¶187.

K. Claim 19

[19.pre]-[19.a]

As discussed for limitations [11.a]-[11.b], [18.d], Aretoulaki-Ramer-Kennewick discloses “*invoking an agent*,” which Kennewick teaches “*process[es]* the commands or questions” corresponding to the natural language utterance. EX1006, [0123]; EX1002, ¶¶188-189.

[19.b]

Aretoulaki’s Knowledge Manager receives the interpretation of the utterance (“domain semantic representation”) and executes a “database query” to retrieve matches from an “application database,” which may be a “*local database*” that the Knowledge Manager “has direct access to.” EX1004, 14:38-40, 15:44-49, 24:12-13; EX1002, ¶190. Furthermore, in the combination where Aretoulaki’s Knowledge Manager is modified to invoke an agent as taught in Kennewick, the agent “create[s] queries to local databases 102 or *though [a] network interface* 116 to data sources on the Internet 42 or other networks” (e.g., a “*remote database*”). EX1006, [0123]; EX1002, ¶191.

L. Claim 20

Aretoulaki discloses maintaining a “user model” (“*user-specific profile*”) containing the user’s “*preferences*.” EX1004, 9:47-49. In the combination, Kennewick discloses providing a “*user interface*” allowing users to modify user-

specific “*preferences*” and/or “manually enter[.]” “user specific *parameters*,” like “presentation format for results.” EX1006, [0073], [0142]-[0143]; EX1002, ¶192. A POSITA would have been motivated to include Kennewick’s user interface in Aretoulaki’s system so users could provide and/or modify their profile to ensure that Aretoulaki’s “user model” includes the user’s latest “preferences.” EX1004, 9:44-52; EX1006, [0031], [0073]; EX1002, ¶193. The combination uses a known technique (user interface to input preferences) to improve a similar system in the same way (enabling a user to update a user model of its preferences through the interface). EX1002, ¶193.

M. Claims 28-29 and 33-42

Claims 28-29 and 33-42 are substantively identical to claims 6-7 and 11-20, respectively, and are obvious for the same reasons. EX1002, ¶¶194-205.

VIII. GROUND 4: KENNEWICK AND HAO

A. Claim 1

[1.pre]

Kennewick discloses methods for processing “*natural language utterances that include requests*,” e.g., “natural language speech questions or...commands.” EX1006, [0002], [0014], [0018]; EX1002, ¶210. Kennewick further discloses “*selecting and presenting purchase opportunities based*” on the processed natural language utterances, by selecting and “provid[ing] interactive offers and promotions

for goods and services” (“*purchase opportunities*”) responsive to user queries and “facilitat[ing] remote ordering.” EX1006, [0060], [0066]-[0067]; EX1002, ¶¶210-211.

Kennewick’s methods are implemented by an “interactive natural language speech system” including a “main speech processing unit” (or “main unit”) that processes utterances received via a “speech unit.” EX1006, [0012], [0121]-[0123]; EX1002, ¶¶212-213. These units are “embedded into” a “computer” that, as shown in FIG. 4, includes “one or more processing units” which are “*one or more physical processors.*” EX1006, [0111], [0115], [0122]; EX1002, ¶¶212-213.

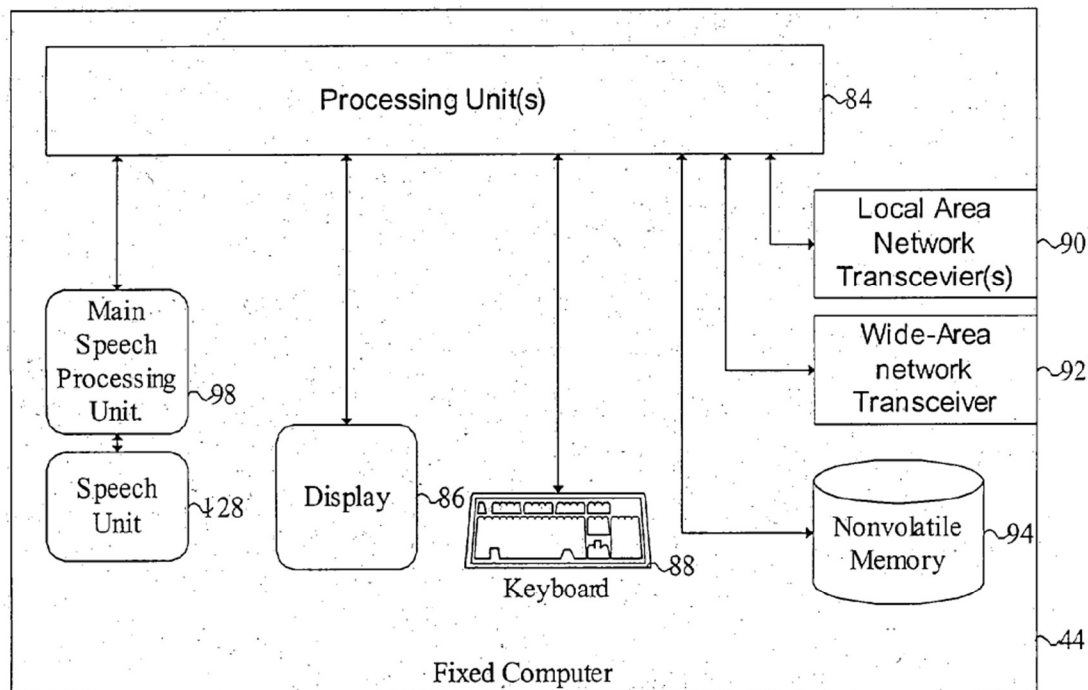


Figure 4. Fixed Computer Block Diagram.

Kennewick's "***one or more physical processors***," which may include the speech unit 128, main speech processing unit 98, and/or processing unit(s) 84, contain various software modules, e.g., "speech recognition module," "parser," and "agents" that are "packages of executable code." EX1006, [0013], [0017]; EX1002, ¶214. These software modules and executable code packages are "***computer program instructions***" that, when executed by the processors of Kennewick's system, cause them to perform the method steps. EX1002, ¶214.

To the extent Kennewick does not explicitly describe how it selects "interactive offers and promotions" ("***purchase opportunities***") based on processed utterances, a POSITA would have found it obvious to look to and incorporate the specific teachings of Hao. EX1002, ¶¶206-209, 215-219. Hao discloses an "e-commerce shopping method" that receives and processes "shopper free form input representing a shopping request" conveyed verbally "in a natural language," meeting the claimed "***natural language utterances that include requests***." EX1007, Abstract, [0009], [0022], claim 8; EX1002, ¶¶215-216. Hao further discloses "***selecting and presenting purchase opportunities based thereon***" by "extract[ing] semantics from" the request and "automatically add[ing]" an item "to an electronic shopping cart" based thereon for presentation to the user, where the user is provided the opportunity to "finalize...an e-commerce transaction" to purchase the presented item(s). EX1007, [0011]-[0012], [0040]; EX1002, ¶¶215-216. A POSITA would

have been motivated to combine Kennewick, which leaves open its method for selecting “interactive offers and promotions,” with Hao, which provides one method of providing those purchase opportunities. EX1002, ¶219. The combination is obvious because it uses a known technique (Hao’s method for selecting purchase opportunities) to improve a similar system (Kennewick, which leaves open the method for selecting purchase opportunities) in the same way. *Id.*, ¶¶206-209, 219.

[1.a]

As shown in FIG. 5, Kennewick’s main unit corresponding to “*one or more processors*” “receives spoken natural language...utterances” from a “speech unit” and provides the utterance as an input to a “*speech recognition engine*” 120. EX1006, [0014], [0149], [0155]; EX1002, ¶220.

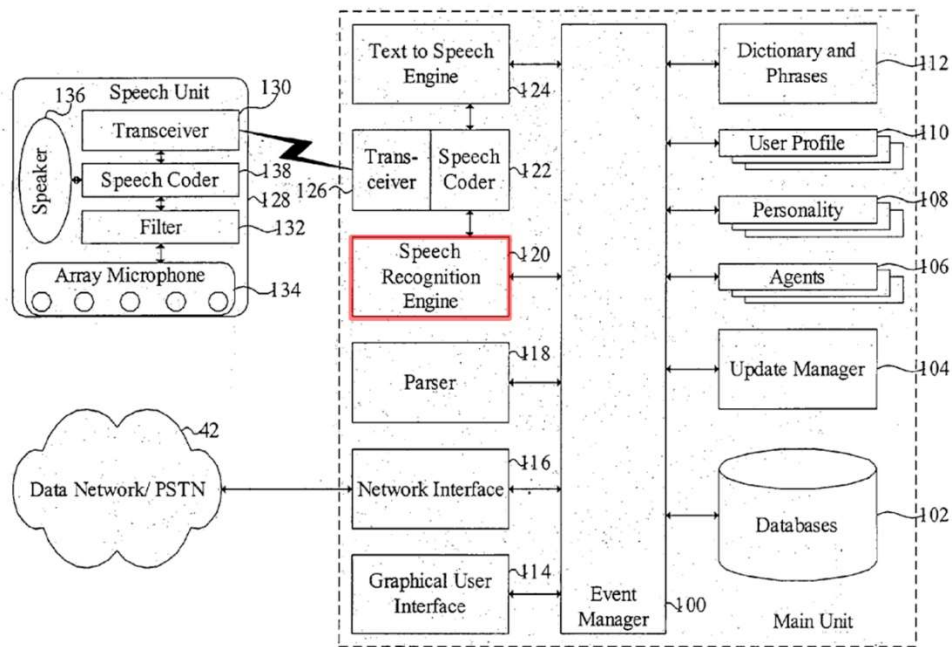


Figure 5. Speech Processing System Block Diagram

[1.b]

Kennewick's speech recognition engine generates and outputs "recognized words and phrases" "*recognized from the natural language utterance.*" EX1006, [0123], [0155]; EX1002, ¶221. The speech recognition engine sends the words and phrases to event manager 100, which sends them to other components such as "the parser 118 for interpretation." EX1006, [0155], [0119], claim 2; EX1002, ¶221. Thus, the main unit "*receiv[es]*" the recognized words or phrases as an output of the speech recognition at the event manager, which it passes to the parser. *Id.*

[1.c]

Kennewick's main unit includes a parser that determines "*a context for the natural language utterance*" based on the recognized words or phrases, by "applying prior probabilities or fuzzy possibilities to keyword matching" performed on "tokens" corresponding to the recognized words or phrases. EX1006, [0032], [0160]-[0161]; EX1002, ¶¶222-224. For example, the parser may determine, based on the recognized words including the keyword "temperature," that the utterance is associated with a weather context or a measurement context, and "use a scoring system to determine" which is most likely based on additional factors such as "the user profile 110, the domain agent's 156 knowledge and previous context." EX1006, [0160]-[0161]; EX1002, ¶222. This is consistent with the '758 patent's disclosure of determining context "by having domain agents 135 competitively

generate a context-based interpretation of the utterance,” and selecting “a highest scoring interpretation.” EX1001, 5:4-14, claim 6; EX1002, ¶223. The ’758 patent further incorporates by reference the disclosure of U.S. Patent No. 7,640,160 for generating “context-based interpretations,” which explains that a parser “determines a context for an utterance” using language identical to Kennewick’s. EX1001, 5:14-25; EX1014, 20:63-21:2; EX1002, ¶223.

[1.d]

Kennewick’s main unit determines the context of the utterance, and, based on the context, “determine[s] the domain and thereby, the domain agent,...to be invoked.” EX1006, [0160]; EX1002, ¶225. The invoked domain agent corresponds to a service providing a respective “area of functionality,” including services enabling the user to purchase fast food, travel services, and/or other goods/services through respective agents. EX1006, [0018], [0060]-[0069]; EX1002, ¶225. Through these services, the user “quer[ies] the system...for goods and services,” and the system responds by “provid[ing] interactive offers and promotions” corresponding to a selected “*purchase opportunity*.” EX1006, [0066]-[0067]; EX1002, ¶225. The system selects the “*purchase opportunity based on the determined context*” corresponding to one of these e-commerce services, e.g., a remote ordering domain. EX1002, ¶225.

To the extent Kennewick does not explicitly disclose how to select the purchase opportunity based on the determined context, the combination with Hao renders it obvious. EX1002, ¶226. As discussed above, Hao discloses “extract[ing] semantics” having “meaning in an e-commerce context” from a received “shopping request,” and matching the “extracted semantics” with “[a]vailable items...specified in an inventory data store” to select “one or more shopping items” that are loaded into a shopping cart meeting the claimed “*purchase opportunity*.” EX1007, [0032]-[0033], [0038]-[0040], [0043], [0045]; EX1002, ¶226. The system selects purchase opportunity “*based on the determined context*” being an e-commerce context, as it analyzes the semantics of the request in the e-commerce context to select the items. See EX1007, [0043]; EX1002, ¶226.

A POSITA would have found it obvious to apply Hao’s teaching of selecting a purchase opportunity in a determined e-commerce context to improve how Kennewick’s system selects and presents “offers and promotions for goods and services.” EX1002, ¶¶227-230. In the combination, Kennewick’s system would process a received natural language utterance to determine a context as discussed above, and invoke a corresponding domain agent managing “the behavior and data” “in [the] specific domain.” EX1006, [0017]-[0018], [0160]; EX1002, ¶¶227-229. For example, if the determined context is an “e-commerce context” corresponding to a particular domain (e.g., remote ordering, fast food ordering, travel services etc.),

a POSITA would have found it obvious for the invoked agent to process the utterance “in [the determined] e-commerce context” to extract semantics and compare them “to inventory items” in an “inventory data store” as taught by Hao to select a “*purchase opportunity based on the determined context.*” *See, e.g.,* EX1007, [0022], [0032]-[0033], [0038]-[0040], [0043], [0045]-[0046]; EX1002, ¶¶227-229; *see also* EX1006, [0123]-[0124] (domain agents generally “create queries to...data sources” to obtain “results” for the user).

A POSITA would have been motivated to make this combination to provide a straightforward way for Kennewick’s system to select a purchase opportunity based on a determined e-commerce context. EX1002, ¶230. A POSITA would have had a reasonable expectation of success because the combination is a simple combination of prior art elements (establishing a context as taught in Kennewick, and selecting a purchase opportunity based on an established context as taught in Hao) according to known methods to yield predictable results (Kennewick’s system establishing a context of an utterance and selecting a purchase opportunity based on the established context). *Id.*, ¶230. A POSITA would also have been motivated to make this combination, which applies a known technique (e.g., Hao’s selecting a purchase opportunity and resolving ambiguities) to improve a similar system (Kennewick’s system through which users can perform remoting ordering) in the

same way, to create an improved interface able to automatically analyze the user's request to select and present a purchase opportunity. *Id.*

[1.e]

Kennewick's main unit delivers a response via a "speech unit 128" for output by a speaker, and/or a "graphical user interface" having a "display 18," which as shown in FIG. 1, are "*electronic device[s] in communication with*" the main speech processing unit 98. EX1006, [0121], [0124]-[0125], [0145]; EX1002, ¶231.

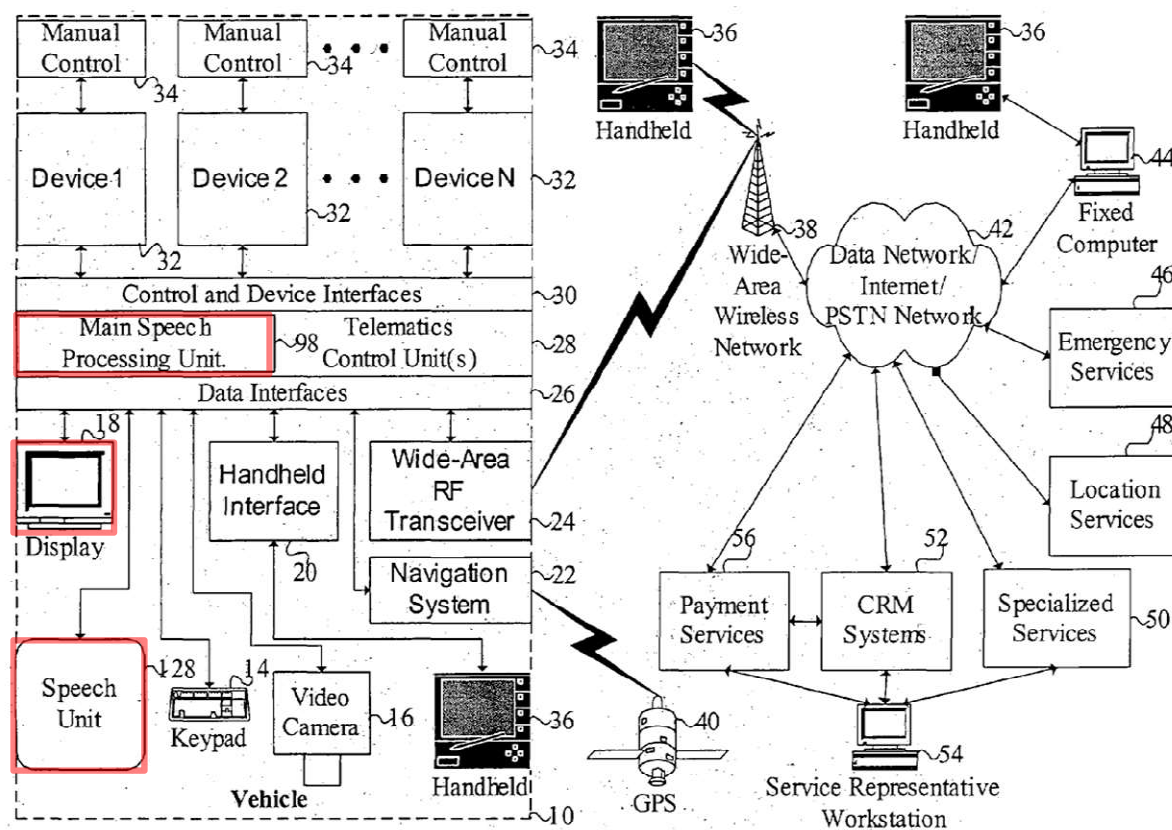


Figure 1. First Embodiment System Block Diagram

The delivered response may include the "*selected purchase opportunity*," e.g., "promotional offers" and/or "product catalog[s]," or "shopper specific shopping

cart...populated with items” in the Kennewick-Hao combination. *See* EX1006, [0066]-[0067]; EX1007, [0040]; EX1002, ¶¶232-234.

[1.f]

Kennewick’s system tracks user interactions with the system to build “user profiles and histories” that are “added [to] and updated automatically as the user uses the system.” EX1006, [0016], [0031], [0143]; EX1002, ¶235. A POSITA would have found it obvious for Kennewick’s tracked interactions to include “***an interaction pattern associated with the [delivered] purchase opportunity***” because Kennewick discloses tracking “customer order history” to facilitate “remote ordering.” EX1006, [0031], [0066]-[0067]; EX1002, ¶236. A POSITA would have understood that a user’s past purchases/ordering corresponds to an “***interaction pattern***” associated with a (past) purchase opportunity, because the user ordering a good/service associated with the purchase opportunity would include them interacting with the system by making “***a subsequent request***” via the electronic device (e.g., speech unit, display, and/or keyboard) that “***interacts with the purchase opportunity***,” e.g., the presented promotion, offer, or shopping cart. *See* EX1006, [0112], [0115], [0125]; EX1002, ¶¶236-237. The subsequent request results in the electronic device “***completing a transaction related to the purchase opportunity***” because Kennewick teaches completing a “remote ordering” transaction using the user’s “secure payment wallet.” *See* EX1006, [0067]; EX1002, ¶237.

In addition, in the Kennewick-Hao combination, Hao teaches maintaining a “shopper profile/preferences data store” storing records of the user’s “past transaction history.” EX1007, [0009], [0027], [0033], [0041]; EX1002, ¶¶238-239. The user interacts with a delivered purchase opportunity, e.g., shopping cart, by making “*a subsequent request*” through a client device (“*electronic device*”) to “finalize”/“*complet[e] a transaction related to the purchase opportunity*,” which the system tracks to maintain the user’s “past transaction history.” EX1007, [0034], [0040]-[0041]; EX1002, ¶238.

A POSITA would have found it obvious for the Kennewick-Hao combination to track interaction patterns associated with delivered purchase opportunities by recording the user’s “past transaction history” as taught by Hao. *See* EX1007, [0027]; EX1002, ¶¶208-209, 239. For instance, Kennewick teaches making “significant use of...user specific profile data” to interpret received utterances, because “some questions” are ambiguous and “can only be interpreted in the context of...the user’s history of interests and preferences.” EX1006, [0007], [0011], [0118]; EX1002, ¶¶208, 239. Kennewick further discloses that each domain “collect[s], store[s] and use[s] specific profile information, as required,” which in e-commerce contexts may include “order history.” EX1006, [0031], [0066]-[0067]; EX1002, ¶¶239. Similarly, Hao discloses tracking a shopper’s “past transaction history...to increase accuracy of future transactions,” and describes specific techniques for

interpreting and resolving ambiguities in a shopper's current request "using shopper specific information" to select items for a purchase opportunity. EX1007, [0033], [0039]-[0041], [0046]; EX1002, ¶¶208, 239. Thus, in the combination where Kennewick is modified to select and deliver purchase opportunities as taught by Hao, a POSITA would have found it obvious to also incorporate Hao's known technique of tracking interaction patterns to maintain a record of "past transaction history" "used to increase accuracy of future transactions," to improve Kennewick's system in the same way. EX1007, [0041]; EX1002, ¶¶208-209, 239.

[1.g]

As shown in FIG. 1, Kennewick's main unit builds and maintains a "***user-specific profile***" (user profile 110) containing the user's "past behavior or history." EX1006, [0016], [0031], [0067], [0143], FIG. 1; EX1002, ¶240.

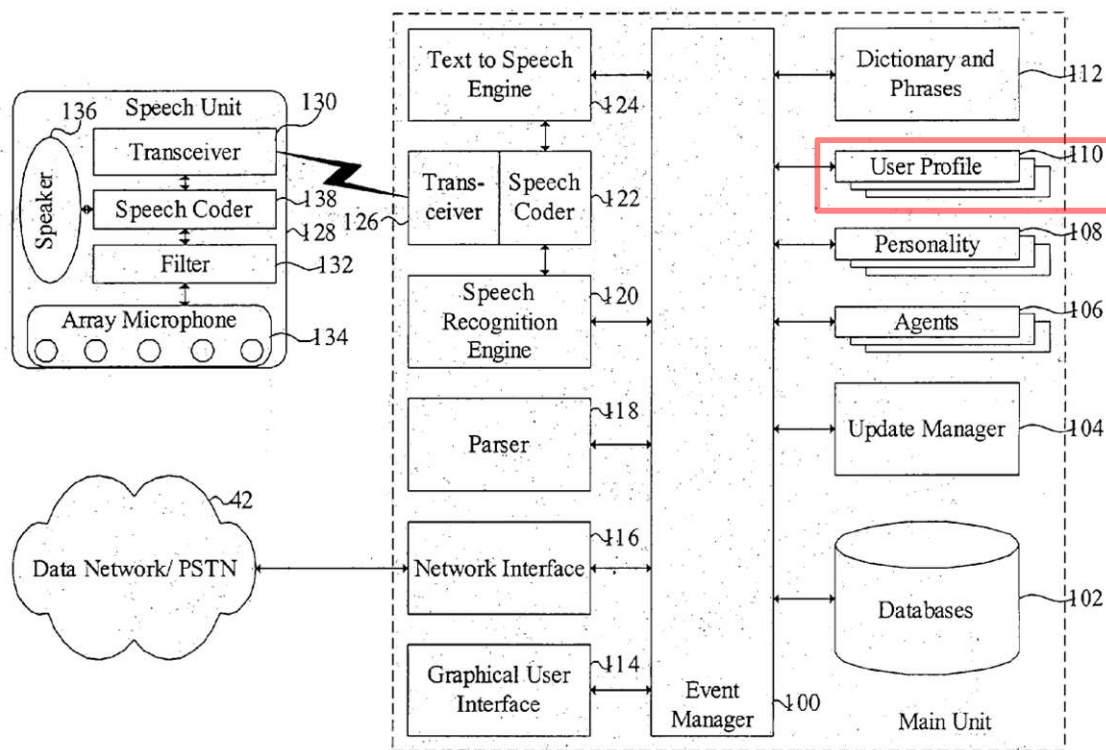
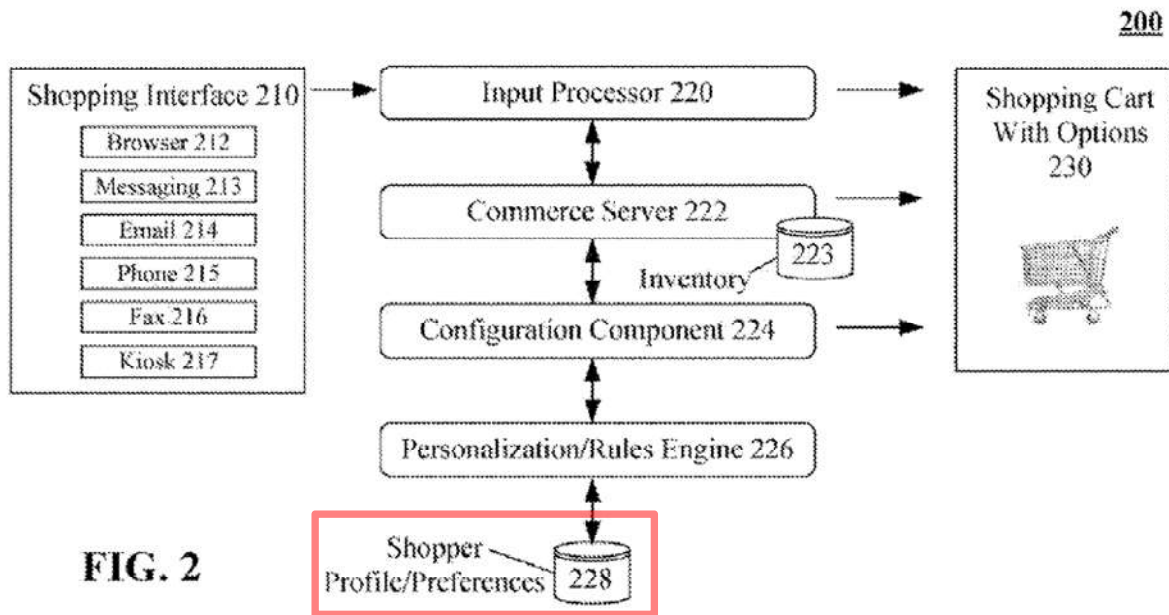


Figure 5. Speech Processing System Block Diagram

Given Kennewick’s disclosure of using “customer order history” to “facilitate[] remote ordering,” a POSITA would have found it obvious for Kennewick’s “*user-specific profile*” to include “order history,” which as discussed above, corresponds to “*tracked interaction pattern[s]*” describing interactions with previously-presented purchase opportunities See EX1006, [0031], [0067], [0143]; EX1002, ¶241. In addition, a POSITA would have found it obvious that when the user makes a new purchase associated with a purchase opportunity, the system automatically “*build[s] or update[es]*” the user-specific profile based on the tracked

interaction pattern to keep user's "order history" up-to-date. See EX1006, [0031]; EX1002, ¶241.

In addition, in the Kennewick-Hao combination, Hao discloses storing "shopper specific information...within a shopper profile/preferences data store" ("*user-specific profile*"), shown in FIG. 2 as "Shopper Profile/Preferences" 228, and updating the profile using "automatically determined shopper data...and historical e-commerce transaction data" ("*tracked interaction pattern[s]*"). EX1007, [0033], [0041], FIG. 2; EX1002, ¶242.



A POSITA would have found it obvious in the combination to build and update Kennewick's "user[-specific] profiles and histories" based on Hao's disclosure of tracking interaction patterns corresponding to "historical e-commerce

transaction data”/“past transaction history.” EX1002, ¶¶208-209, 243. A POSITA would have been motivated to make this combination to maintain and use the user’s “historical e-commerce transaction data” as taught by Hao to “facilitate[] remote ordering,” by enabling the system to select and present purchase opportunities even when the user’s request contains ambiguities. EX1006, [0067]; EX1007, [0032]-[0033], [0041]; EX1002, ¶¶208-209, 243. A POSITA would have found this combination obvious as a combination of prior art elements (e.g., Kennewick’s “user profiles and histories,” and Hao’s “shopper specific information” including “historical e-commerce transaction data”) according to known methods to yield predictable results. EX1002, ¶243.

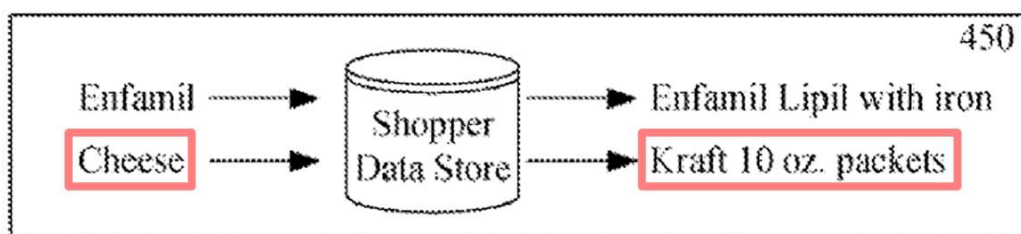
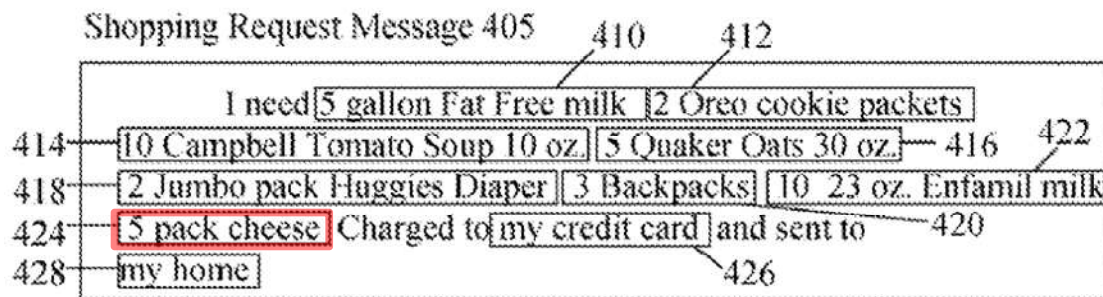
[1.h]

Kennewick’s main unit “*interpret[s]*” received natural language utterances using “user profile information” (“*user-specific profile*”), and selects “offers and promotions for goods and services” (“*purchase opportunity*”) based thereon. EX1006, [0028], [0031]-[0032], [0066]-[0067], [0123], [0143]; *see also* limitation [1.d]; EX1002, ¶244. These include processing a “*subsequent natural language utterance*” using user profile information built from interactions involving previous utterances and purchase opportunities, to select a “*subsequent purchase opportunity*.” EX1002, ¶244.

To the extent Kennewick does not explicitly disclose how to “*interpret[]...[the] subsequent natural language utterance*” and “*select [the] subsequent purchase opportunity*” as claimed, the Kennewick-Hao combination renders it obvious. EX1002, ¶¶245-251. As discussed above, Hao uses “past transaction history” from a “shopper profile” (“*user-specific profile*”) to “increase accuracy of future transactions.” EX1007, [0033], [0041]; EX1002, ¶245. For example, when interpreting a user’s request, Hao discloses performing a “check for ambiguities.” EX1007, FIG. 3, [0039]; EX1002, ¶245. If one is found, a “configuration component” “automatically resolve[s]” the ambiguity using the user’s shopper profile, e.g., “purchase history,” to “produce[] results specifying one or more items” for populating a “shopping cart.” EX1007, [0009], [0039], FIG. 3, EX1002, ¶245. The shopping cart items are a “*subsequent purchase opportunity*” selected based on the “*subsequent natural language utterance*” “*interpret[ed]...using the user-specific profile*” to resolve the ambiguity. EX1002, ¶245.

Hao’s FIG. 4, portions of which are reproduced below, provides an example of interpreting an ambiguous user request using the “*user-specific profile*” to select a “*subsequent purchase opportunity*.”

400



EX1007, FIG. 4; EX1002, ¶¶246-247. As shown in FIG. 4, Hao’s input processor analyzes the words or phrases from the request and determines that it includes “ambiguous items,” e.g., “5 packs of cheese.” EX1007, [0042]-[0045]; EX1002, ¶¶247-248. Hao then uses the user’s “purchase history, preference, and profile information” to interpret the “ambiguous item” as referring to “Kraft 10 oz. packets,” and places the identified item into the shopping cart as part of the “*subsequent purchase opportunity*.” EX1007, [0042], [0045]-[0046], FIG. 4; EX1002, ¶¶247-248. While Hao describes FIG. 4 as relating to “free form text processing,” a POSITA would have understood that the example applies equally to a “*subsequent natural language utterance*,” because Hao discloses using an “automatic speech recognition (ASR) engine” to “convert” any received speech utterances “into text”

for further processing “by the natural language understanding engine.” EX1007, [0009]-[0010], [0022], claim 15; EX1002, ¶¶246, 249. Thus, the described “free form text processing” applies equally to text-based inputs and to speech-based inputs converted into text through ASR. EX1002, ¶249.

A POSITA would have found it obvious in the Kennewick-Hao combination to modify Kennewick’s remote ordering services to incorporate Hao’s teaching of interpreting subsequent natural language utterances using a user-specific “shopper profile” to select a subsequent purchase opportunity. EX1002, ¶¶208-209, 250-251. For example, Kennewick teaches that “some questions” are ambiguous and “can only be interpreted in the context of...the user’s history of interests and preferences,” and that “customer preferences” and “customer order history” can be used to facilitate “[r]emote ordering and payment for goods and services.” EX1006, [0007], [0067]; EX1002, ¶¶208, 250. Hao teaches using a shopper’s “past transaction history...to increase accuracy of future transactions,” and describes specific techniques for interpreting received requests and resolving ambiguities “using shopper specific information.” EX1007, [0033], [0040]-[0041]; EX1002, ¶¶208, 250. By resolving these ambiguities, the system automatically selects a purchase opportunity, without needing to solicit additional information from the user. *See* EX1007, [0039]-[0040], [0046]; EX1002, ¶¶208, 250.

A POSITA would have been motivated to incorporate Hao’s techniques for interpreting ambiguous user requests based on user-specific information, e.g., “past transaction history,” to select and present purchase opportunities to the user, to improve how Kennewick uses “user specific profile data” to facilitate remote ordering, by enabling Kennewick’s system to interpret user requests to select a purchase opportunity even when the request is “incomplete, ambiguous or subjective,” thereby “minimiz[ing] the interaction with the user required to deliver a response.” EX1007, [0039], [0041], [0045]-[0047]; EX1006, [0002], [0007], [0032], [0067]; EX1002, ¶¶209, 251. This modification is a combination of prior art elements disclosed in Kennewick and Hao (e.g., Kennewick’s use of “customer order history” to facilitate remote ordering, and Hao’s specific techniques of using “past transaction history” to resolve ambiguities and select purchase opportunities) according to known methods to yield predictable results. EX1002, ¶¶209, 251.

B. Claim 2

[2.pre]

Kennewick-Hao discloses that “the purchase opportunity” may relate to “*an item*,” e.g., “items to purchase,” or “at least one *item* available for sale.” EX1006, [0070]; EX1007, [0032]-[0033]; EX1002, ¶252.

[2.a]

Kennewick's main unit (e.g., “*one or more processors*”) receives natural language utterances, which may include a “*second natural language utterance.*” *See, e.g.,* EX1006, [0121]-[0123], [0155], [0177] (describing interpreting “subsequent commands” based on “previous commands”); EX1002, ¶253.

[2.b]

In the Kennewick-Hao combination, when a purchase opportunity is presented to the user, the user can make “*a request to purchase the item*” by requesting to “finalize...an e-commerce transaction involving a purchase of” present item(s) as taught by Hao, to perform “[r]emote ordering and payment for goods.” EX1007, [0011], [0041]; EX1006, [0067]; EX1002, ¶254. Because Kennewick and Hao disclose that the user may interact with the system through an interactive speech interface that processes the user's natural language utterances, a POSITA would have found it obvious for this request to be in the form of a natural language utterance, e.g., a “*second natural language utterance.*” *See* EX1006, [0101]; EX1007, claim 20 (“voice only interface”); EX1002, ¶255. In the combination, Kennewick's parser interprets the received utterance and “transforms [it] into complete commands and questions,” which a POSITA would have found obvious may be a command/request to “finalize” the “e-commerce transaction” to “*purchase*

the item,” enabling Kennewick’s remote ordering service to complete the remote ordering transaction. *See* EX1006, [0067], [0123]; EX1002, ¶256.

A POSITA would have found it obvious to modify Kennewick, to the extent not explicitly disclosed, to enable users to select an option to finalize an e-commerce transaction to purchase the item as taught by Hao, as it provides a straightforward way for Kennewick’s system to perform remote ordering. EX1002, ¶256. A POSITA would have found such a modification to be obvious, as it simply involves applying a known technique (the user requesting to purchase an item associated with a presented purchase opportunity) to a known device (Kennewick’s speech interface enabling remote ordering) ready for improvement to yield predictable results (the user performing remote ordering by requesting to purchase an item associated with a presented purchase opportunity). *Id.*

[2.c]

Kennewick “facilitates remote ordering” by enabling the user to, through the “natural language speech interface,” complete a purchase transaction (“ordering and payment”) for goods, e.g., an item. *See* EX1006, [0067], [0101]; EX1002, ¶257. Similarly, Hao discloses that the user may use voice commands to “finalize the e-commerce transaction” to purchase an item. *See* EX1007, [0041], claim 20; EX1002, ¶257. Thus, Kennewick-Hao discloses the claimed “*completing...a*

purchase transaction of the item” based on the received voice command (“*second natural language utterance*”) to finalize the transaction. EX1002, ¶257.

C. Claim 4

Kennewick-Hao discloses delivering the selected purchase opportunity via the electronic device. EX1002, ¶¶258-259. Given Kennewick’s disclosure of presenting responses in an audible, conversational (e.g., not “rigidly formatted or mechanical”) manner “*by the electronic device*” (i.e., the speech unit 128 having speaker 136), a POSITA would have found it obvious for the purchase opportunity to be presented “*audibly*” as a “*natural language conversational purchase opportunity*.” See, e.g., EX1006, [0024], [0121], [0124]; see also EX1007, claim 20 (“voice only interface”); EX1002, ¶¶258-259.

D. Claim 5

[5.pre]

The processors of Kennewick’s system contain software modules corresponding to the claimed “*computer program instructions*,” including, as shown in FIG. 1, a “parser” and “agents” which operate together to “*interpret a natural language utterance*.”

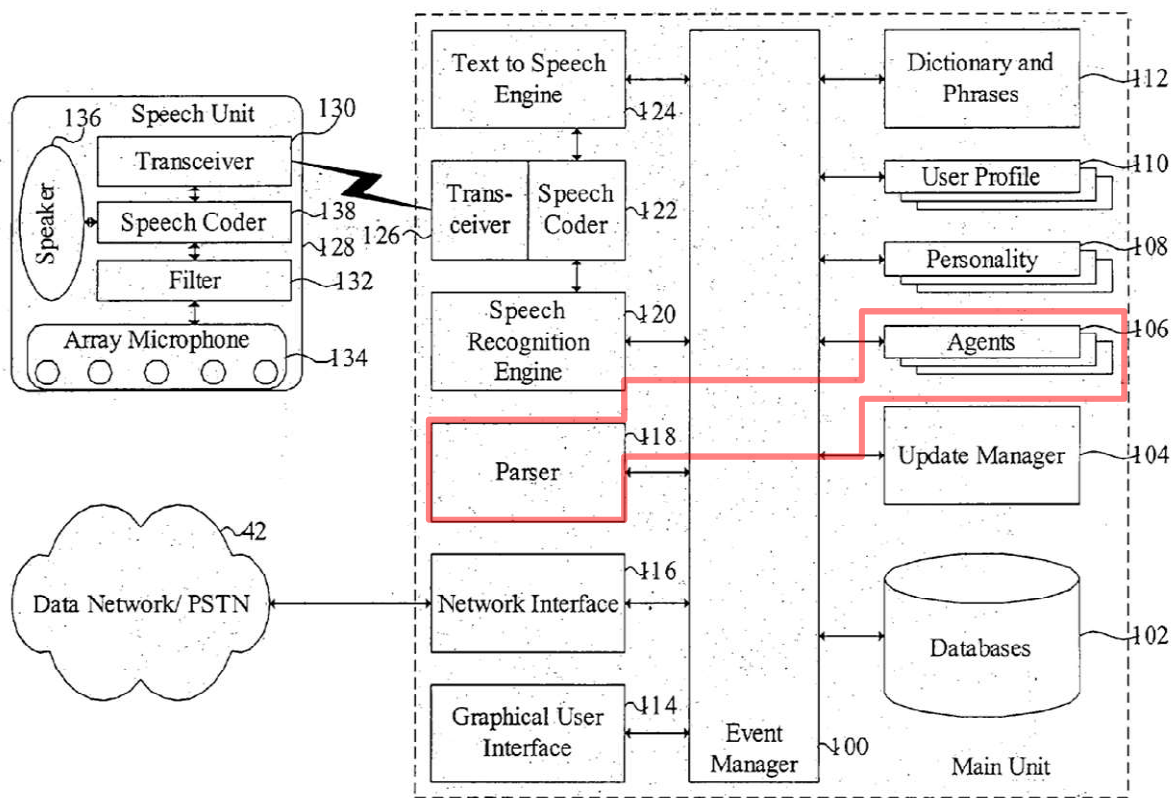


Figure 5. Speech Processing System Block Diagram

EX1006, [0013], [0017]; FIG. 1; EX1002, ¶¶260-261. For example, the parser interprets the natural language utterance by “determin[ing] a context for the utterance” and “transform[ing] [it] into complete commands and questions using data supplied by the agents.” EX1006, [0123], [0155], [0160]-[0163]; *see also* EX1001, 4:31-37, 5:9-14, 14:35-40, FIG. 1 (showing “conversational language processor” as including “context determination module” and “agents”); EX1002, ¶¶261-262.

[5.a]

As discussed for limitation [1.b], Kennewick's main unit receives the “**words or phrases...as an output of the speech recognition engine,**” e.g., “speech recognition module,” at the “event manager 100.” See EX1006, [0119], [0123], [0155]; EX1002, ¶263. The main unit then “pass[es]”, or “**provid[es]**,” these words or phrases as input “to the parser 118,” which is part of the “**conversational language processor,**” “for interpretation.” EX1006, [0119], [0155]; EX1002, ¶263.

[5.b]

Kennewick's parser and agents (“**conversational language processor**”) “pars[e] and interpret[]” each utterance by “transform[ing]” the words and phrases of the utterance “into complete commands and questions” corresponding to the “**interpretation of the natural language utterance.**” EX1006, [0028], [0123], [0155], [0162]-[0163]; EX1002, ¶264. The main unit “**receives**” the interpretation as “**an output of the conversational language processor,**” e.g., as an output of the parser 118 via the event manager 100, and passes it to an invoked domain agent, which uses the interpretation to query “one or more...data sources” to obtain results to be presented to the user as a response. See EX1006, [0028], [0119], [0127], [0165]; EX1002, ¶¶264-265. A POSITA would have found it obvious that in contexts where the utterance is a “query...for offers and promotions for goods and services,” the main unit invokes a domain agent which selects the “**purchase**

opportunity” (e.g., “offers and promotions”) based on the queries generated using the interpretation of the utterance. *See* EX1006, [0066]; EX1002, ¶265.

To the extent that Kennewick does not explicitly disclose selecting “***the purchase opportunity***” as claimed, the combination with Hao renders it obvious. As discussed above, Hao discloses interpreting the utterance by extracting “semantic meanings” from the recognized text of the utterance, and selecting “***the purchase opportunity***” based on “***the interpretation***” by comparing the extracted semantic meanings to available “inventory items.” EX1007, [0022], [0028], [0043]; EX1002, ¶¶266-267. In the combined system, a POSITA would have found it obvious to modify Kennewick’s system to select the purchase opportunity based on the interpretation as taught by Hao. EX1002, ¶267. A POSITA would have recognized the combination as applying a known technique (e.g., Hao’s selection of a purchase opportunity by comparing the interpretation to available items) to improve a similar system (e.g., Kennewick’s system which presents purchase opportunities in response to user queries) in the same way, so that Kennewick’s system compares the interpretation of the natural language utterance to available items to select the purchase opportunity. EX1002, ¶267. A POSITA would have had a reasonable expectation of success, as the combination is a combination of prior art elements disclosed in Kennewick and Hao according to known methods to yield predictable results. *Id.*

E. Claim 6

Kennewick's “*conversational language processor*” includes agents, which as shown in FIG. 6, include “*one or more domain agents.*” EX1006, [0018], [0160]; EX1002, ¶268.

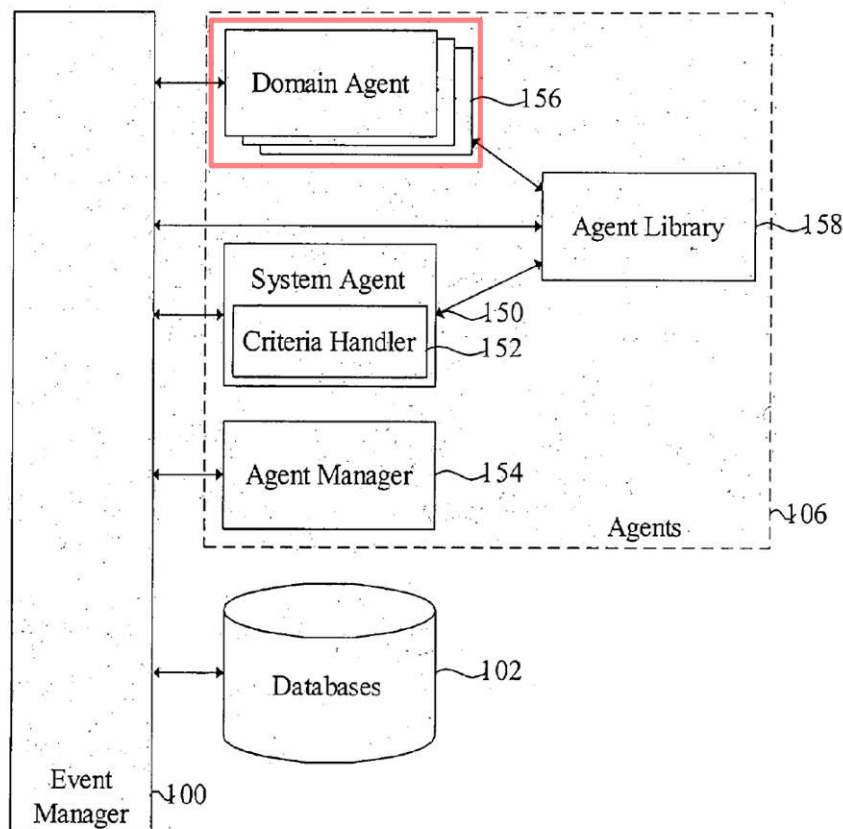


Figure 6. Agent Architecture

Each domain agent provides “behavior and data required for a specific area of functionality” to assist the parser in “*interpreting requests related to its domain,*” e.g., by “suppl[ying]” domain-specific information for the parser “to determine the most plausible interpretation” of an utterance. EX1006, [0018], [0029], [0123],

[0163]; EX1002, ¶269. Each domain agent further assists in “**determining a response to the request relating to its domain**” by “creat[ing] one or more queries to...information sources” related to its domain, for “creat[ing] a satisfactory response to” the request. EX1006, [0167], [0178]; EX1002, ¶270.

As discussed above, a POSITA would have found it obvious to modify Kennewick such that when the invoked domain agent relates to an e-commerce service, e.g., “fast food ordering agent” or “remote ordering” agent, it assists in “**determining a response to the [user’s] request relating to [the invoked] domain**” by selecting the purchase opportunity as taught by Hao. See EX1007, [0022]; EX1006, [0018], [0066]-[0067]; EX1002, ¶271. A POSITA would have found this modification obvious, as each domain agent may have its own “domain specific behavior,” and it would be obvious for a domain agent for an e-commerce service to determine “offers or promotions” for goods/services responsive to user requests using the teachings of Hao to select the goods/services, as discussed for claim 5. See EX1006, [0017], [0066]; EX1002, ¶271.

F. Claim 7

Kennewick’s “**one or more domain agents**” assist in “specific domains,” including **electronic commerce domain[s]**” for travel services, entertainment scheduling, and remote ordering and payment for goods such as fast food. EX1006,

[0018], [0067], [0127]; EX1002, ¶¶272-273; *see also* EX1007, Abstract (selecting purchase opportunities via “an e-commerce shopping method”).

G. Claim 9

Kennewick’s system can be deployed on “multiple devices...placed at different locations,” where the system uses data of the locations of each device, e.g., an “*environmental model*,” to determine the “location of the particular device addressed by the user,” e.g., “*environmental information*,” which it uses “as part of the context for” the natural language utterance. EX1006, [0016]; EX1002, ¶274; *see also* EX1001, claim 10. Alternatively, Kennewick’s system maintains “user profile and histories,” which are “*an environmental model*” from which it determines “*environmental information*” corresponding to past user activity and actions. *See* EX1006, [0031]; EX1002, ¶275; *see also* EX1001, claim 10. The system determines “*the context for the natural language utterance*” based on the user’s past activity and actions. EX1006, [0160]; EX1002, ¶275.

H. Claim 10

Kennewick discloses that the environmental information may comprise “the location of the particular device addressed by the user” (“*user location*”) and/or a “history of [the user’s] interactions with the system” (“*user activity*” or “*user action*”). EX1006, [0016], [0143], [0160]; EX1002, ¶276.

I. Claim 11

[11.pre]-[11.a]

Kennewick's system processes each user query by "invoking the proper resources, including agents" which are "***system resources***" available to the system, which in the combination, includes invoking "the correct agent" based on the determined context to "***to select the purchase opportunity based on the determined context***" based on the teachings of Hao. EX1006, [0017], [0028]; EX1002, ¶¶277-280.

[11.b]

The invoked "***system resources***" include "agents" ("***an agent***"), including a "domain agent," which Kennewick discloses is "an autonomous executable" corresponding to a "re-distributable package[]...of functionality...for a specific domain" or "application area," that "receives, processes and responds to" received "user questions, queries, and[/or] commands." EX1006, claim 1, [0017], [0028], [0126], [0160]; EX1002, ¶281.

J. Claim 12

Kennewick's "agent" meets the claimed "***agent***" because it "return[s] results of[user] questions as responses to users," where the responses are created "using the results of information queries, the system personality 108[,] and the user preferences

or other data in the user[-specific] profile 110.” EX1006, [0124]; EX1002, ¶¶282-283.

K. Claim 13

Kennewick discloses an “update manager,” shown in FIG. 5, for “**updating the agent**” “remotely over a network” to “add[] new behaviors,” which a POSITA would have understood corresponds to “**new functionality**.” EX1006, [0017], [0098], [0120], [0126]; *see also id.*, [0129] (“[d]evelopers of agents can add new functionality to the agent library”); EX1002, ¶¶284-285.

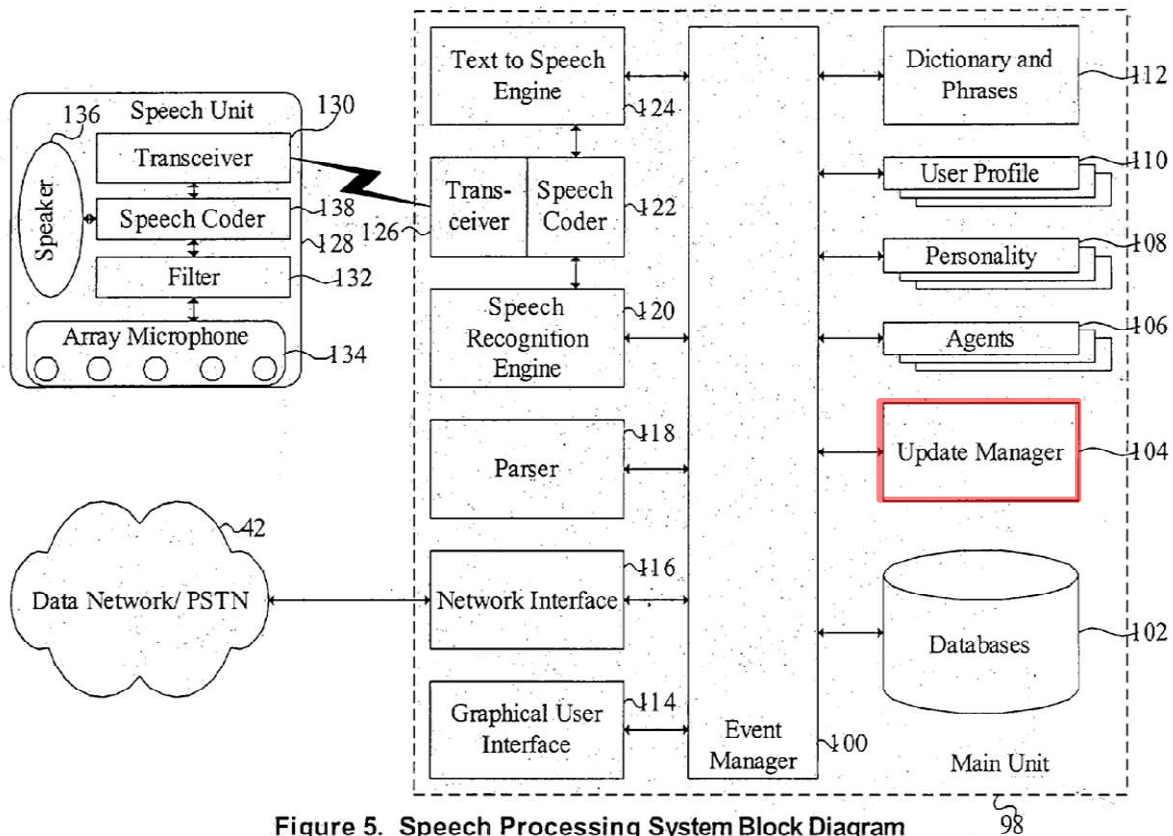


Figure 5. Speech Processing System Block Diagram

98

L. Claim 14

Kennewick discloses that its “agents,” including the claimed “*agent*,” meets the claim because they use “system resources and the services of other, typically more specialized, agents.” EX1006, [0017]; EX1002, ¶286.

M. Claim 15

Kennewick discloses providing a “*framework*” such as speech or graphical user interface allowing the user to “create new agents” and/or “modify...existing agents.” EX1006, [0134]; EX1002, ¶287.

N. Claim 16

Kennewick discloses “*providing a license management framework*” by providing “license management capability allowing the sale of agents by third parties to one or more users *on a one time or subscription basis.*” EX1006, [0126], [0132]; EX1002, ¶288.

O. Claim 17

[17.pre]-[17.a]

Kennewick’s system implementing the method “can be deployed in *a network of devices using [a] common base of agents, data, information, user[-specific] profiles and histories.*” EX1006, [0016], claim 27; EX1002, ¶¶289-290. The user can “*interact with, and receive the same services and applications* at any location equipped with the required device on the network,” or in other words, from “*any one of the devices on the network.*” *Id.*

[17.b]

As discussed above, Kennewick's system is "deployed in *a network of devices*...placed at different locations," and can "*use the location of [a] particular device addressed by the user as part of the context for the questions asked*" of the user. EX1006, [0016]; EX1002, ¶291.

P. Claim 18

[18.pre]-[18.a]

Kennewick's system includes "a speech recognition module for *capturing* the user input" in the form of "spoken natural language...utterances from a user," including a "*first natural language utterance.*" EX1006, [0013], [0014], [0028]; EX1002, ¶¶292-293.

[18.b]

Kennewick's system includes "a parser for *parsing* the input," e.g., the "*first natural language utterance.*" EX1006, [0013], [0028]; EX1002, ¶294.

[18.c]

Kennewick's system performs "parsing and *interpreting* [of] the [user's] question or query" e.g., "*first natural language utterance.*" EX1006, [0028]; EX1002, ¶295.

[18.d]

Kennewick discloses invoking "the required agent" to process and respond to the user's request, selected from a library of agents 106. EX1006, [0028], [0032],

[0127]-[0128]; EX1002, ¶¶296-297. For example, Kennewick’s FIG. 6 illustrates an “agent architecture” that includes a collection or library of domain agents managed by an “agent manager 154.”

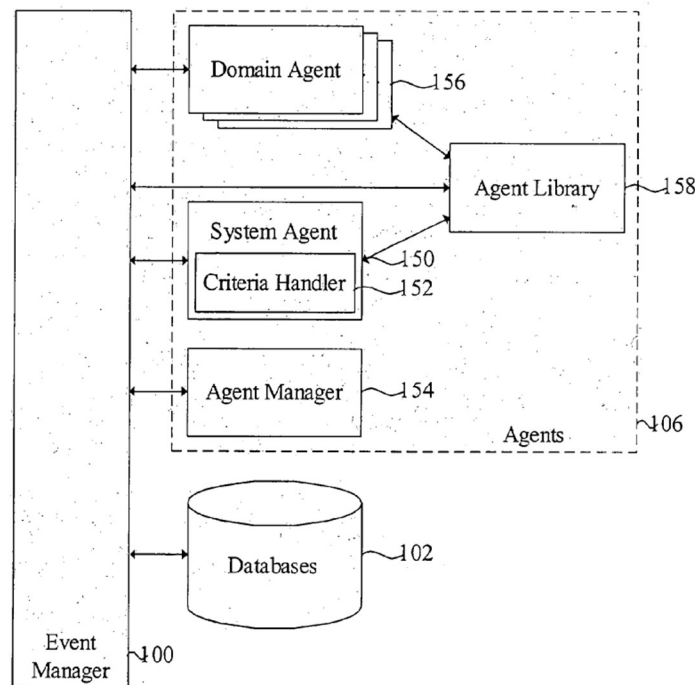


Figure 6. Agent Architecture

See EX1006, [0126], FIG. 6; EX1002, ¶297. A POSITA would have understood that Kennewick’s “required agent” is “*an applicable agent from a library of agents*,” because a POSITA would have found it obvious that a managed collection of domain agents may be a “library” of agents. EX1002, ¶¶296-297.

Q. Claim 19

[19.pre]-[19.a]

As discussed for limitations [11.a]-[11.b], [18.d], Kennewick-Hao discloses that the invoked agent “*process/es* the commands or questions” corresponding to the natural language utterance to select the purchase opportunity. EX1006, [0123]; EX1002, ¶¶298-299.

[19.b]

Kennewick’s invoked agent processes the natural language utterance by “create[ing] queries to *[a] local database[]* 102 or though [a] network interface 116 to data sources on the Internet 42 or other networks,” e.g., a “*remote database.*” EX1006, [0123]; EX1002, ¶300.

R. Claim 20

Kennewick provides a “*user interface*” allowing users to “modify existing *questions* and “modify existing *questions* and *queries*,” modify user-specific “*preferences*,” and/or “manually enter[]” “user specific *parameters.*” EX1006, [0021], [0073], [0142]-[0143]; EX1002, ¶301.

S. Claim 21

As discussed for limitation [1.c], Kennewick discloses determining “a context”/domain based on the recognized words or phrases through “keyword matching.” EX1006, [0160]; EX1002, ¶302. Kennewick further discloses that the contexts/domains “are broadly divided into two categories, query and response, and

control.” EX1006, [0018]; EX1002, ¶302. Thus, Kennewick discloses that the determined context comprises “*a categorization of the recognized words or phrases*,” because it determines context by matching the recognized words and phrases to the determined context/domain, categorizing them as being associated with “query and response” or “control.” EX1002, ¶302.

T. Claim 22

Kennewick discloses determining “the most likely context or domain,” comprising one of *commands* (e.g., control domain), *directories* (“directories of travel and entertainment services”), *e-commerce* (“[r]emote ordering and payment”), *education* (“[i]nteractive educational programs”), *entertainment* (“entertainment scheduling”), *sports* (“[s]ports scores”), *subscriptions* (subscriptions “to multimedia information channels”), *transportation* (“driving directions” and/or “travel services”), or *weather* (“[w]eather information”), etc. EX1006, [0018], [0043]-[0051], [0060]-[0072], [0074]-[0085]; EX1002, ¶303.

U. Claim 23

Claim 23 is rendered obvious for similar reasons as claim 1. EX1002, ¶304. The claimed “*system*” is met by Kennewick-Hao’s “interactive natural language speech system” which performs “*processing [of] natural language utterances that include requests and selecting and presenting purchase opportunities based thereon*,” and include “*one or more physical processors programmed with*

computer instructions, which when executed cause the one or more physical processors to” perform the claimed steps. *See* EX1006, [0012], [0095]-[0097], [0111], [0115], claim 1; EX1002, ¶304. The remaining elements of claim 23 are substantively identical to those of claim 1, and are obvious for the same reasons. EX1002, ¶304.

V. Claims 24, 26-29, 31-34

Claims 24, 26-29, and 31-34 are substantively identical to claims 2, 4-7 and 9-22, respectively, and are obvious for the same reasons. EX1002, ¶¶305-323.

IX. GROUND 5: KENNEWICK, HAO, AND ARETOULAKI

A. Claim 3

[3.pre]/[3.a]

In Kennewick-Hao, Hao teaches that the user can “*interact[] with the purchase opportunity*,” e.g., to change or purchase the presented item(s). EX1007, [0011]; EX1002, ¶¶324-325. The system receives “*an indication of an occurrence*” of this interaction. *Id.*

In addition, a POSITA would have been motivated to modify the combination based on Aretoulaki, which as discussed above, discloses monitoring/tracking user interactions with the purchase opportunity to “locate communication or understanding problems” and “trigger a repair mechanism.” EX1004, 7:10-24, 18:19-25, 32:21-25; EX1002, ¶326. Aretoulaki thus discloses receiving “*an*

indication of an occurrence of a user interaction with the purchase opportunity”

that is part of the tracked interaction pattern, because Aretoulaki discloses “constantly monitoring”/tracking the user’s interactions with the system, including interactions with any presented purchase opportunities. EX1002, ¶¶326-327.

A POSITA would have been motivated to combine Aretoulaki with Kennewick-Hao, as the references are analogous, all relating to speech-based interfaces through which users search for and purchase products, and which track user interactions with presented purchase opportunities. *See, e.g.*, EX1006, Abstract, [0007], [0018], [0066]-[0068], [0143]; EX1007, Abstract, [0009], [0027], [0011], [0033], [0040]; EX1004, Abstract, 8:27-48, 9:47-49, 11:31-36, 22:10-23; EX1002, ¶328. A POSITA in would have been motivated modify Kennewick-Hao to, after presenting a purchase opportunity, “constantly monitor” user interactions as taught by Aretoulaki, enabling the user “to correct [or interrupt] the system...in the case of a misunderstanding,” allowing the system to achieve Kennewick’s goal of being able to achieve “graceful recovery” after “full or partial failure.” EX1004, 7:11-41; EX1006, [0011]; EX1002, ¶329. A POSITA would have had a reasonable expectation of success because it combines prior art elements of tracking user interactions according to known methods to yield the predictable result of Kennewick-Hao being able to track additional types of interactions as taught by

Aretoulaki to detect and recover from misunderstandings. *See, e.g.*, EX1004, 9:22-35; EX1006, [0118]; EX1002, ¶330.

[3.b]

In the Kennewick-Hao-Aretoulaki combination discussed above, Aretoulaki discloses determining “*whether the interpretation of the natural language utterance was correct based on the indication*,” by determining whether “*the indication*” corresponds to a “Correct” dialog act indicating that the user “intends to correct the instantiation for a parameter,...because the system has misrecognized or misinterpreted a previous user input,” e.g., the “*natural language utterance*.” EX1004, 12:30-35, 38:44-51; EX1002, ¶¶331-332. A POSITA would have found it obvious to modify Kennewick-Hao to implement Aretoulaki’s repair mechanisms to correct misunderstandings. EX1002, ¶333. For example, Kennewick teaches making “significant use of context [and] prior information” to interpret user utterances, and it would be obvious to further use this information to determine whether a user interaction represents a “break[] in the normal flow” signifying a misunderstanding/misinterpretation as taught in Aretoulaki. *See* EX1006, [0011], [0118]; EX1004, 18:8-25; EX1002, ¶333. This would enable a more robust natural language interface able to locate and repair “communication or understanding problems,” to recover from “full or partial failure.” EX1006, [0011]; EX1002, ¶333. A POSITA would have had a reasonable expectation of success because the

modification is a combination of known prior art elements (e.g., Kennewick’s use of context and prior information to interpret utterances, and Aretoulaki’s use of tracked interactions and prior information to detect misinterpretations), combined according to known methods to yield predictable results (e.g., Kennewick’s system using tracked interactions and prior information to determine when the system has misinterpreted the user’s request). *See, e.g.*, EX1006, [0118]; EX1004, 9:22-35, 18:8-25; EX1002, ¶334.

[3.c]

In the Kennewick-Hao-Aretoulaki combination, the system interprets natural language utterances using “context [and] prior information,” which a POSITA would have understood includes the prior information of “*whether the interpretation of the [previous] natural language utterance was correct.*” EX1006, Abstract, [0011]; EX1002, ¶¶335-336. For example, Aretoulaki discloses that a “repair mechanism” updates the system’s interpretation of an utterance by establishing a “new/modified parameter value” when a previous interpretation was incorrect. EX1004, 15:37-42, 18:19-25, 39:1-5; EX1002, ¶336. Afterwards, the system interprets subsequent natural language utterances “*based on the determination of whether the interpretation of the natural language utterance was correct,*” by interpreting them “in terms of the preceding exchanges between the user and the system,” including the “new/updated parameter value” set by the repair

mechanism. EX1004, 4:17-23, *see also id.*, 16:38-40, 22:28-31; EX1001, 5:28-32; EX1002, ¶336.

A POSITA would have understood that in the combination incorporating Aretoulaki's teachings of monitoring user interactions with the purchase opportunity to determine whether the interpretation of the natural language utterance was correct, the system interprets the subsequent natural language utterance in the same way as discussed above. EX1002, ¶337. A POSITA would have found this combination obvious, as it aligns with Kennewick's teachings of making "significant use of context [and] prior information" when interpreting received utterances, and achieving "graceful recovery" from prior failures. EX1004, [0011]; EX1002, ¶337. The modification is a combination of prior art elements disclosed in Kennewick and Aretoulaki (e.g., Kennewick's system using tracked interactions and prior information to interpret subsequent utterances, and Aretoulaki's repair mechanism updating a previous interpretation used in interpreting subsequent utterances) according to known methods to yield predictable results. EX1002, ¶337.

B. Claim 25

Claim 25 is substantively identical to claim 3, and is obvious for the same reasons. EX1002, ¶338.

X. GROUND 6: KENNEWICK, HAO, AND JONG

A. Claim 8

To the extent Kennewick-Hao does not explicitly disclose recognizing the words or phrases using “*phonetic dictation*,” the combination with Jong renders it obvious. EX1002, ¶¶339-340. While Kennewick-Hao leaves open the specific speech recognition technique used, Jong, as discussed above, discloses a specific speech recognition technique using “*phonetic dictation*” that translates “speech input signals into textual representations” of a stream of phonemes, and recognizes “*words or phrases...based on*” “word syllables” extracted from “*the recognized stream of phonemes*.” EX1008, Abstract, 5:50-67, 9:34-46; EX1002, ¶340. A POSITA would have been motivated to modify Kennewick-Hao’s speech recognition by looking to Jong’s specific techniques as a straightforward way to implement Kennewick-Hao’s speech recognition. EX1002, ¶¶341-342. A POSITA would have had a reasonable expectation of success because the combination is a simple substitution of Jong’s specific known speech recognition techniques for Kennewick-Hao’s more general disclosure to yield predictable results, e.g., Kennewick-Hao’s speech recognition including phonetic dictation to recognize phonemes as taught in Jong. *Id.*, ¶342.

B. Claim 30

Claim 30 is substantively identical to claim 8 and is obvious for the same reasons. EX1002, ¶343.

XI. CONCLUSION

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

XII. STANDING

Petitioner certifies that Petitioner is not barred or estopped from requesting this *inter partes* review and the '758 patent is IPR-eligible.

XIII. MANDATORY NOTICES AND FEES

A. Real Party-in-Interest

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

B. Related Matters

Patent Owner asserts the '758 patent against Petitioner in *VB Assets, LLC v. Amazon.com Services, LLC*, No. 1:24-cv-00839 (District of Delaware, July 18, 2024) (the “parallel litigation”).

Petitioner is aware of the following additional related matters involving the '758 patent and/or related patents:

Petition for *Inter Partes* Review
U.S. Patent No. 11,080,758

Case Caption	Forum	Patents
<i>VB Assets LLC v. SoundHound AI Inc.</i> , Case No. 1:24-cv-01279	DDE	7,818,176 10,755,699 11,087,385 11,222,626 10,297,249 9,269,097 9,502,025 8,886,536 8,073,681
<i>VB Assets LLC v. Amazon.com Services LLC et al.</i> , Case No. 1:19-cv-01410	DDE	7,818,176 9,626,703 9,015,049 9,269,097 8,886,536 8,073,681
<i>VB Assets LLC v. Apple Inc.</i> , Case No. 1:24-cv-01368	DDE	7,818,176 8,073,681 10,297,249 10,510,341 8,515,765 8,886,536 9,269,097
<i>VB Assets LLC v. Samsung Electronics Co., Ltd. et al</i> , Case No. 2:24-cv-00828	E.D. Tex.	8,073,681 10,755,699 10,510,341 8,515,765 8,886,536 7,818,176
<i>VB Assets LLC v. Amazon.com Services LLC</i> , Case No. 25-1142	Fed. Cir.	7,818,176 9,626,703 9,015,049 9,269,097 8,886,536 8,073,681

Case Caption	Forum	Patents
<i>VB Assets LLC v. Amazon.com Services LLC</i> , Case No. 25-1357	Fed. Cir.	7,818,176 8,073,681 9,269,097
<i>Amazon.com Services LLC v. VB Assets LLC</i> , Case No. 25-1113	Fed. Cir.	7,818,176 9,626,703 8,073,681 9,269,097
IPR2020-01390	PTAB	7,818,176
IPR2025-00867	PTAB	7,818,176
IPR2020-01388	PTAB	9,269,097
IPR2020-01374	PTAB	8,886,536
IPR2020-01377	PTAB	8,886,536
IPR2025-00869	PTAB	8,886,536

C. Lead and Backup Counsel

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

D. Service Information

Petitioner consents to service by electronic mail at:

VBAssets-IPR@fenwick.com.

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

E. Fees

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

Dated: July 1, 2025

FENWICK & WEST LLP

/J. David Hadden/

J. David Hadden
Reg. No. 40,629

Attorneys for Petitioner
Amazon.com Services LLC

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. §42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. § 42.8, certificates of service or word count, or appendix of exhibits, contains 13,925 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: July 1, 2025

FENWICK & WEST LLP

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

Attorneys for Petitioner
Amazon.com Services LLC

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

I hereby certify, pursuant to 37 C.F.R. §§ 42.6(e) and 42.105, that a complete copy of this Petition for *Inter Partes* Review including all exhibits (Nos. 1001-1017) are being served via Federal Express on July 1, 2025, upon Patent Owner by serving the correspondence address of record with the USPTO as follows:

James Gatto
Daniel Yannuzzi
David Heisey
Sheppard Mullin Richter & Hampton
650 Town Center Drive, 10th Floor
Costa Mesa, CA 92626
*Patent Owner's correspondence
address of record for US Patent 11,080,758*

Dated: July 1, 2025

FENWICK & WEST LLP

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

Attorneys for Petitioner
Amazon.com Services LLC