

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

DECLARATION OF PADHRAIC SMYTH, PH.D.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	19
II. BACKGROUND AND QUALIFICATIONS	19
III. MATERIALS CONSIDERED	23
IV. BASIS OF OPINION	26
A. Legal Standard for Anticipation	26
B. Legal Standard for Obviousness	27
C. Level of Ordinary Skill in the Art	29
D. Claim Construction	31
V. OVERVIEW OF THE '758 PATENT	32
VI. RELATED LITIGATION AND RELATED PATENTS	34
VII. PRIOR ART TO THE CHALLENGED PATENT	35
A. Technical Background	35
B. Summary of the Prior Art	45
1. Aretoulaki (EX1004)	45
2. Ramer (EX1005)	49
3. Kennewick (EX1006)	53
4. Hao (EX1007)	55
5. Jong (EX1008)	59
VIII. SUMMARY OF INVALIDITY GROUNDS	60
IX. INVALIDITY GROUNDS	60
A. Ground 1: Aretoulaki in combination with Ramer render obvious claims 1-5, 9-10, 21-27, 31-32, 43-44	60

TABLE OF CONTENTS

(Continued)

	Page
1. General Motivations to Combine Aretoulaki and Ramer	60
2. Claim 1	65
[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:”	65
[1.a]: “providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;”	69
[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;”	71
[1.c] “determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;”	72
[1.d] “selecting, by the one or more processors, a purchase opportunity based on the determined context;”	74
[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;”	75

TABLE OF CONTENTS (Continued)

	Page
[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;”	77
[1.g] “building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and”	85
[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.”	89
3. Claim 2	95
[2.pre] “The method of claim 1, wherein the purchase opportunity relates to an item, the method further comprising:”	95
[2.a] “receiving, by the one or more processors, a second natural language utterance;”	96
[2.b] “determining, by the one or more processors, that the second natural language utterance comprises a request to purchase the item; and”	97

TABLE OF CONTENTS

(Continued)

	Page
[2.c] “completing, by the one or more processors, a purchase transaction of the item based on the second natural language utterance.”	98
4. Claim 3	101
[3.pre] The method of claim 1, the method further comprising:”	101
[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;”	102
[3.b] “determining, by the one or more processors, whether the interpretation of the natural language utterance was correct based on the indication; and”	103
[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.”	104
5. Claim 4: “The method of claim 1, wherein delivering the selected purchase opportunity via the electronic device comprises: audibly presenting, by the electronic device, a natural language conversational purchase opportunity.”	105
6. Claim 5	106
[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:”	106

TABLE OF CONTENTS (Continued)

	Page
[5.a] “providing, by the one or more processors, the words or phrases as an input to the conversational language processor; and”	108
[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.”	109
7. Claim 9: “The method of claim 1, the method further comprising: 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.”	111
8. Claim 10: “The method of claim 1, wherein the environmental information comprises a user location, a user activity, or a user action.”	115
9. Claim 21: “The method of claim 1, wherein the context comprises a categorization of the recognized words or phrases.”	115
10. Claim 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.	116
11. Claim 23	117
12. Claim 24	117
13. Claim 25	117

TABLE OF CONTENTS (Continued)

	Page
14. Claim 26	118
15. Claim 27	118
16. Claim 31	118
17. Claim 32	118
18. Claim 43	118
19. Claim 44	118
 B. Ground 2: Aretoulaki in combination with Ramer and Jong render obvious claims 8 and 30	 119
1. Claim 8: “The method of claim 1, wherein recognizing the words or phrases from the natural language utterance comprises: 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.”	 119
2. Claim 30	122
 C. Ground 3: Aretoulaki in combination with Ramer and Kennewick render obvious claims 6-7, 11-20, 28-29, and 33- 42	 123
1. General Motivations to Combine Aretoulaki, Ramer, and Kennewick.....	123
2. Claim 6: “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: (i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.”	 129

TABLE OF CONTENTS (Continued)

	Page
3. Claim 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.”	134
4. Claim 11	136
[11.pre] “The method of claim 1, further comprising:”	136
[11.a] “invoking, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;”	136
[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.”	138
5. Claim 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.”	138
6. Claim 13: “The method of claim 11, further comprising updating the agent from a remote location over a network to provide new functionality.”	141
7. Claim 14: “The method of claim 11, wherein the agent uses system resources and the services of other agents.”	142
8. Claim 15: “The method of claim 11, further comprising providing a framework allowing the user to create or modify the agent.”	143

TABLE OF CONTENTS (Continued)

	Page
9. Claim 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.”	143
10. Claim 17	144
[17.pre] “The method of claim 11, wherein:”	144
[17.a] “the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories the user can interact with, and receive the same services and applications from, any one of the devices on the network;”	144
[17.b] “the location of a particular device addressed by the user is used as part of context for questions asked of the user.”	148
11. Claim 18	149
[18.pre] “The method of claim 1, further comprising:”	149
[18.a] “capturing, by the one or more processors, the first natural language utterance;”	149
[18.b] “parsing, by the one or more processors, the first natural language utterance;”	150
[18.c] “interpreting, by the one or more processors, the first natural language utterance; and”	151
[18.d] “invoking, by the one or more processors, an applicable agent from a library of agents.”	153
12. Claim 19	156
[19.pre] “The method of claim 1, further comprising:”	156

TABLE OF CONTENTS

(Continued)

	Page
[19.a] “invoking an agent to process the natural language utterance; and”	156
[19.b] “creating a query to a local database or through a network interface to a remote database.”	157
13. Claim 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.”	158
14. Claim 28	160
15. Claim 29	160
16. Claim 33	160
17. Claim 34	161
18. Claim 35	161
19. Claim 36	161
20. Claim 37	161
21. Claim 38	161
22. Claim 39	161
23. Claim 40	161
24. Claim 41	162
25. Claim 42	162
D. Ground 4: Kennewick in combination with Hao render obvious claims 1-2, 4-7, 9-24, 26-29, and 31-44	162
1. General Motivations to Combine Kennewick and Hao	162
2. Claim 1	167

TABLE OF CONTENTS

(Continued)

	Page
[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:”	167
[1.a]: “providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;”	176
[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;”	177
[1.c] “determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;”	178
[1.d] “selecting, by the one or more processors, a purchase opportunity based on the determined context;”	181
[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;”	188

TABLE OF CONTENTS

(Continued)

	Page
[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;”	192
[1.g] “building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and”	196
[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.”	200
3. Claim 2	210
[2.pre] “The method of claim 1, wherein the purchase opportunity relates to an item, the method further comprising:”	210
[2.a] “receiving, by the one or more processors, a second natural language utterance;”	211
[2.b] “determining, by the one or more processors, that the second natural language utterance comprises a request to purchase the item; and”	211

TABLE OF CONTENTS

(Continued)

	Page
[2.c] “completing, by the one or more processors, a purchase transaction of the item based on the second natural language utterance.”	213
4. Claim 4: “The method of claim 1, wherein delivering the selected purchase opportunity via the electronic device comprises: audibly presenting, by the electronic device, a natural language conversational purchase opportunity.”	214
5. Claim 5	216
[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:”	216
[5.a] “providing, by the one or more processors, the words or phrases as an input to the conversational language processor; and”	218
[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.”	219
6. Claim 6: “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: (i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.”	223

TABLE OF CONTENTS

(Continued)

	Page
7. Claim 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.”	226
8. Claim 9: “The method of claim 1, the method further comprising: 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”	228
9. Claim 10: “The method of claim 1, wherein the environmental information comprises a user location, a user activity, or a user action.”	230
10. Claim 11	231
[11.pre] “The method of claim 1, further comprising:”	231
[11.a] “invoking, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;”	231
[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.”	234
11. Claim 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.”	235

TABLE OF CONTENTS (Continued)

	Page
12. Claim 13: “The method of claim 11, further comprising updating the agent from a remote location over a network to provide new functionality.”	236
13. Claim 14: “The method of claim 11, wherein the agent uses system resources and the services of other agents.”	238
14. Claim 15: “The method of claim 11, further comprising providing a framework allowing the user to create or modify the agent.”	238
15. Claim 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.”	238
16. Claim 17	239
[17.pre] “The method of claim 11, wherein:”	239
[17.a] “the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories the user can interact with, and receive the same services and applications from, any one of the devices on the network; and”	239
[17.b] “the location of a particular device addressed by the user is used as part of context for questions asked of the user.”	240
17. Claim 18	240
[18.pre] “The method of claim 1, further comprising:”	240
[18.a] “capturing, by the one or more processors, the first natural language utterance;”	240

TABLE OF CONTENTS

(Continued)

	Page
[18.b] “parsing, by the one or more processors, the first natural language utterance;”	241
[18.c] “interpreting, by the one or more processors, the first natural language utterance; and”	241
[18.d] “invoking, by the one or more processors, an applicable agent from a library of agents.”	242
18. Claim 19	243
[19.pre] “The method of claim 1, further comprising:”	243
[19.a] “invoking an agent to process the natural language utterance; and”	244
[19.b] “creating a query to a local database or through a network interface to a remote database.”	244
19. Claim 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.”	245
20. Claim 21: “The method of claim 1, wherein the context comprises a categorization of the recognized words or phrases.”	246
21. Claim 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.	246
22. Claim 23	247
23. Claim 24	248
24. Claim 26	248

TABLE OF CONTENTS **(Continued)**

	Page
25. Claim 27	248
26. Claim 28	248
27. Claim 29	249
28. Claim 31	249
29. Claim 32	249
30. Claim 33	249
31. Claim 34	249
32. Claim 35	249
33. Claim 36	249
34. Claim 37	250
35. Claim 38	250
36. Claim 39	250
37. Claim 40	250
38. Claim 41	250
39. Claim 42	250
40. Claim 43	250
41. Claim 44	251
E. Ground 5: Kennewick in combination with Hao and Aretoulaki render obvious claims 3 and 25.....	251
1. Claim 3	251
[3.pre] The method of claim 1, the method further comprising:”	251

TABLE OF CONTENTS (Continued)

	Page
[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;”	251
[3.b] “determining, by the one or more processors, whether the interpretation of the natural language utterance was correct based on the indication; and”	256
[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.”	258
2. Claim 25	261
F. Ground 6: Kennewick in combination with Hao and Jong render obvious claims 8 and 30	261
1. Claim 8: “The method of claim 1, wherein recognizing the words or phrases from the natural language utterance comprises: 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”	261
2. Claim 30	264
X. CONCLUSION.....	264

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

I. INTRODUCTION

1. I have been asked by the party requesting this review, Amazon.com Services LLC (“Petitioner”), to provide my expert opinions in support of its petition for *inter partes* review of U.S. Patent No. 11,080,758 (the “’758 patent” or “challenged patent”) challenging the patentability of claims 1-44 (the “challenged claims”).

2. In summary, it is my opinion that all challenged claims are unpatentable for the reasons I provide in this declaration.

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

II. BACKGROUND AND QUALIFICATIONS

4. The details of my education, work experience, research, patents, and publications are summarized in my CV (EX1017). I provide here a summary of some of the relevant information from my CV, as well as other relevant background information.

5. I earned a Ph.D. and Master’s degree in Electrical Engineering from the California Institute of Technology. I earned a Bachelor of Engineering (Electronic)

from the National University of Ireland, University College Galway in 1984. My senior undergraduate thesis was on speech recognition.

6. At present, I am a Distinguished Professor and Endowed Chair in Artificial Intelligence in the Department of Computer Science at the University of California, Irvine. Prior to that, I was a Member of Technical Staff and Technical Group Leader at the Jet Propulsion Laboratory.

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

8. During my career I have worked extensively on research related to key aspects of automated speech recognition, natural language processing, user modeling, and human dialog. I have attached my CV (EX1017), which provides more details about the experience I summarize here. As an example, since 1992 I have conducted a significant body of research and published a series of papers on the topic of hidden Markov models (HMMs). HMMs are a general statistical framework for pattern recognition in sequential data (such as audio waveforms or

text) and were a key component in automated speech recognition systems and natural language understanding since the 1980's. My research on HMMs produced significant advances in the understanding of the statistical and algorithmic foundations of speech recognition and natural language understanding. These advances included development of new HMM methods for detecting anomalous patterns in time-series signals, new algorithms for automated clustering of waveforms using HMMs, new classes of HMMs with applications to speech modeling, and novel HMM methods for automated segmentation of waveforms. My papers on the topic of HMMs have been collectively cited over 1,800 times in the literature.

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

ontologies, and analysis of large language models. As with my work on HMMs, my work in natural language understanding meant that I was, and am, broadly familiar with the research literature in this area.

10. I have also conducted research in user modeling, focusing on statistical and machine learning techniques for automated generation of user models from observational data. In addition, I have published papers on techniques for automated analysis of human dialog, such as automatically classifying utterances into specific topics. As I discuss below, user modeling is an important function of many interactive conversational systems, which enables such systems to manage a dialog with the user and infer information about the user's requests.

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

12. I have also served in editorial roles and conference chair positions for major journals and conferences in my research areas, including serving as a founding editorial board member for the Journal of Machine Learning Research, an associate editor for the Journal of the American Statistical Association, program chair for the

Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining conference, co-program chair for the International Joint Conference on Artificial Intelligence, and serving as Senior Area Chair for the Neural Information Processing Systems conference and the International Machine Learning Conference on multiple occasions.

13. I am also a technical consultant with private industry in areas including machine learning, user modeling, signal processing, pattern recognition, and natural language understanding. Over the past 25 years, I have consulted in these areas for companies including Toshiba, Microsoft, Google, SAP, Netflix, eBay, AT&T, Samsung, Nokia, as well as several startup companies.

III. MATERIALS CONSIDERED

14. I have reviewed and considered, in preparing this declaration, the '758 patent (EX1001), its prosecution history (EX1003), and the following prior art references:

- WO Pub. No. 02/073331 to Aretoulaki et al., filed February 20, 2002, and titled "Natural Language Context-Sensitive and Knowledge-Based Interaction Environment for Dynamic And Flexible Product, Service and Information Search and Presentation Applications" ("Aretoulaki") (EX1004);

- U.S. Patent No. 7,752,209 to Ramer et al., filed January 19, 2006, and titled “Presenting Sponsored Content on a Mobile Communication Facility” (“Ramer”) (EX1005);
- U.S. Publication No. 2004/0193420 to Kennewick et al., filed July 15, 2003, and titled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance” (“Kennewick”) (EX1006);
- U.S. Publication No. 2008/0189187 to Hao et al., filed February 1, 2007, and titled “Populating an E-commerce Shopping Cart and Other E-commerce Fields Based Upon Content Extracted from Natural Language Input” (“Hao”) (EX1007); and
- U.S. Patent No. 6,173,250 to Jong, filed June 3, 1998, and titled “Apparatus and Method for Speech-Text-Transmit Communication Over Data Networks” (“Jong”) (EX1008).

15. I am also aware of and considered in forming my opinions presented in this declaration, other information available to a person of ordinary skill in the art at the time of the priority date of the challenged patent (“POSITA”), such as other

patents, reference materials, articles, and standards, and knowledge of a POSITA.¹

These include:

- Prosecution History of U.S. Patent No. 11,080,758 (EX1003);
- U.S. Patent No. 7,376,586 to Partovi et al., filed December 17, 1999, and titled “Method and Apparatus for Electronic Commerce Using a Telephone Interface” (“Partovi”) (EX1009);
- 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”) (EX1010);
- Lucente, M., “Conversational interfaces for e-commerce applications,” Communications of the ACM (Sept. 2000), 43(9), 59-61, (“Lucente”) (EX1011);
- Huang et al., “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development,” Prentice Hall (2001), (“SLP”) (EX1012);

¹ I understand that the challenged patent has a priority date of February 6, 2007, as I discuss in more detail below.

- U.S. Patent No. 6,859,776 to Cohen et al., filed October 4, 1999 and titled “Method and Apparatus for Optimizing a Spoken Dialog Between a Person and a Machine” (“Cohen”) (EX1013); and
- U.S. Patent No. 7,640,160 to Di Cristo et al., filed August 5, 2005 and titled “Systems and Methods for Responding to Natural Language Speech Utterance” (“Di Cristo”) (EX1014).

16. Should new information be disclosed or discovered during this proceeding, I may supplement and further elaborate on the opinions I express in this declaration as necessary. Nothing in this declaration should be interpreted as an opinion that no grounds of invalidity exist other than the ones in the petition. I do not express such an opinion and have not reached such an opinion.

IV. BASIS OF OPINION

A. Legal Standard for Anticipation

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

the claim, it may nevertheless anticipate if a POSITA would envisage the claimed arrangement or combination.

B. Legal Standard for Obviousness

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

19. It is my understanding that “obviousness” is a question of law based on underlying factual issues including (1) the scope and content of the prior art, (2) the differences between the prior art and the asserted claims, (3) the level of ordinary skill in the pertinent art, and (4) the existence of secondary considerations such as commercial success, long-felt but unresolved needs, and failure of others.

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

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

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

C. Level of Ordinary Skill in the Art

22. I understand that the '758 patent should be interpreted from the perspective of a person having “ordinary skill in the art” as of the priority date of the patent application (i.e., a POSITA). For the purposes of this declaration, I assume that the priority date of the patent application leading to the '758 patent is no earlier than February 6, 2007, and I understand that Patent Owner has not claimed an earlier priority date in the parallel litigation. I am familiar with the technological field at issue and the state of the art as of that date.

23. I understand that the education level of those in the technological field, sophistication of the technology, types of problems encountered, prior art solutions to those problems, and the speed at which innovations are made in the field may factor into establishing what the appropriate level of ordinary skill should be.

24. The specification of the challenged patent states that it relates to using “natural language models to deliver targeted advertisements” and/or “purchase opportunities,” and more specifically, to selecting advertisements/purchase opportunities based on “user profiles” built from “track[ed] historical data about users.” EX1001, Abstract, 2:65-3:1, 5:65-6:3, 11:12-18, 12:1-9. The claims of the challenged patent are directed to “[a] method for processing natural language utterances that include requests” to “select[] and present[] purchase opportunities.” *Id.*, 12:61-18:24. The steps of the method include providing “a natural language

utterance as an input to a speech recognition engine,” receiving “words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine,” determining “a context for the natural language utterance based on the recognized words or phrases,” selecting “a purchase opportunity based on the determined context,” delivering “the selected purchase opportunity via an electronic device,” tracking “an interaction pattern associated with the purchase opportunity delivered to the electronic device” that “includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity” and “complet[es] a transaction related to the purchase opportunity,” building or updating “a user-specific profile based on the tracked interaction pattern,” and interpreting “a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.” *Id.*, *e.g.*, 12:61-13:32 (claim 1).

25. In my opinion, 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.

26. I had at least the requisite level of skill with respect to the subject matter of the challenged patent at the time of the priority date as demonstrated by my qualifications listed above and my CV (EX1017).

D. Claim Construction

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

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

29. I understand that a district court in *VB Assets, LLC, v. Amazon.com, Inc., et al.*, C.A. No. 19-1410 (MN) (D. Del.) construed the term “model” to mean “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system,” and the term “speech recognition engine” as “software or hardware that recognizes the words or phrases in the natural language utterance,” for U.S. Patent Nos. 7,818,176 (“’176 patent”), 8,886,536 (“’536 patent”), and 9,269,097 (“’097 patent”), which are parent patents of the challenged patent that

share the same specification. EX1016, 1-2. My opinion is that these constructions are consistent with the plain and ordinary meaning of these terms.

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

V. OVERVIEW OF THE '758 PATENT

31. The '758 patent was filed on November 19, 2018, and claims earliest priority to February 6, 2007. I understand that Patent Owner has not claimed an earlier priority date in the district court proceeding. For the purposes of this proceeding, I treat February 6, 2007 as the '758 patent's priority date.

32. The '758 patent is directed to systems and methods “to deliver targeted advertisements” or “purchase opportunities,” “and/or provide natural language processing based on advertisements” or purchase opportunities. EX1001, Abstract, 2:65-3:1.

33. The challenged patent discloses a system 100 that includes an input device 105, an Automatic Speech Recognizer (“ASR”) 110, a conversational language processor 120, and an output device 180. EX1001, Fig. 1 (reproduced below).

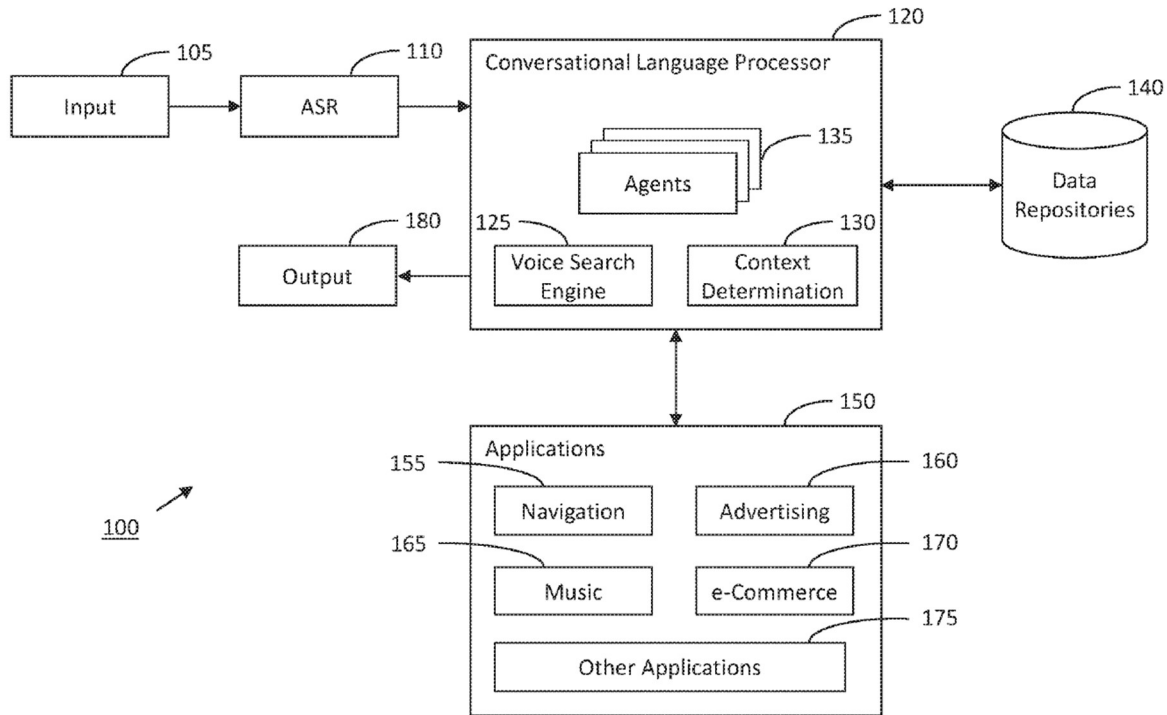


Figure 1

34. The input device 105 receives a user utterance (e.g., at a microphone). *See* EX1001, 4:7-11. The ASR 110 processes the utterance to recognize “words and phrases in [the] utterance.” *Id.*, 3:60-64, 4:14-15. The conversational language processor may then use “context determination module 130 to establish a context . . . by having domain agents 135 competitively generate a context-based interpretation” of the utterance, e.g., “by scoring possible interpretations and selecting a highest scoring interpretation.” *Id.*, 5:4-9. Based on the “context-based interpretation,” the conversational language processor “generate[s] a domain-

specific conversational response” that is “returned to the user as an output.” *Id.*, 4:39-42.

35. The ’758 patent also discloses selecting and presenting a purchase opportunity, e.g., an advertisement, in response to a user’s utterance. EX1001, 11:19-40. The system may “track historical data about users, conversations, topics, contexts” or “other criteria” “to build statistical profiles defining affinities . . . , or other criteria on a user-specific and/or a global-user level.” *Id.*, 12:2-14. The system selects purchase opportunities like advertisements “using various techniques, including content of the request” and “user profiles” or “preferences.” *Id.*, 11:12-18.

VI. RELATED LITIGATION AND RELATED PATENTS

36. I understand that the ’758 patent is the subject of pending litigation involving Petitioner and Patent Owner: *VB Assets, LLC v. Amazon.com Services, LLC*, No. 1:24-cv-00839 (District of Delaware, July 18, 2024) (the “parallel litigation”). I understand that Patent Owner sued Petitioner for alleged infringement of the ’758 patent. *See* EX1015.

37. I also understand that that the ’758 patent is related to the ’176 patent, ’536 patent, and ’097 patent. The ’176 patent was at issue to the following district court litigations and IPRs: *VB Assets LLC v. Apple Inc*, Case No. 1:24-cv-01368 (D. Del.); *VB Assets LLC v. SoundHound AI Inc*, Case No. 1:24-cv-01279 (D. Del.), *VB Assets LLC v. Amazon.com Services LLC et al.*, Case No. 1:19-cv-01410 (D. Del.),

VB Assets LLC v. Samsung Electronics America Inc et al., Case No. 2:24-cv-00828 (E.D. Tex.), IPR2020-01390, and IPR2025-00867. The '536 patent was at issue to the following district court litigations and IPRs: *VB Assets LLC v. Apple Inc*, Case No. 1:24-cv-01368 (D. Del.); *VB Assets LLC v. SoundHound AI Inc*, Case No. 1:24-cv-01279 (D. Del.), *Assets LLC v. Amazon.com Services LLC et al.*, Case No. 1:19-cv-01410 (D. Del.), *VB Assets LLC v. Samsung Electronics America Inc et al.*, Case No. 2:24-cv-00828 (E.D. Tex.), IPR2020-01374, IPR2020-01377, and IPR2025-00869. The '097 patent was at issue to the following district court litigations and IPRs: *VB Assets LLC v. Apple Inc*, Case No. 1:24-cv-01368 (D. Del.); *VB Assets LLC v. SoundHound AI Inc*, Case No. 1:24-cv-01279 (D. Del.), *VB Assets LLC v. Amazon.com Services LLC et al.*, Case No. 1:19-cv-01410 (D. Del.), and IPR2020-01388.

VII. PRIOR ART TO THE CHALLENGED PATENT

A. Technical Background

38. In this section, I discuss the state of the art relevant to the subject matter of the '758 patent. The '758 patent describes ideas related to (i) a human user interacting with a system via a natural language utterance, (ii) the system analyzing the recognized words of the user's speech to determine a context and select a purchase opportunity that is delivered to the user, (iii) the system tracking the user's interaction with the delivered purchase opportunity to build or update a user-specific

profile, and (iv) the system using the user-specific profile to interpret a subsequent natural language utterance from the user and to select a subsequent purchase opportunity. These ideas were well-known in the prior art at the time of the '758 patent.

39. In interactive conversational systems, involving a human user and a system, a dialog (or conversation) can be represented as a series of utterances, where each utterance is alternately generated by the human and by the system. A simple example of a dialog might be the following: “User: I would like to fly to New York. System: When would you like to travel? User: This Friday please.” System: “What is your departure airport?” and so on. In this example there are two utterances by the user and two by the system. Technologies and systems to broadly support such interactive conversations had been under development and were well-known prior to 2006, both in academic research laboratories and in industry. *See, e.g.*, EX1010; EX1011; EX1012, Chapters 17-18. These speech-based applications included voice-only interfaces, where the conversation is conducted entirely in the speech modality, as well as multi-modal interfaces, in which speech could be combined with other types of inputs such as typed text, mouse clicks, or touches. For example, interactive voice response (IVR) systems implemented over telephony interfaces were typically speech-only. *See* EX1012, Chapter 18. On the other hand, speech applications implemented on devices with display functionality could take

advantage of the complementary strengths of different input modalities. *See* EX1011; EX1012, Chapter 18.2.1.3.

40. These types of interactive conversational systems, as developed and implemented by 2007, generally relied on a number of related technologies that I will briefly describe below. The primary technologies are automated speech recognition (ASR), natural language understanding (NLU), dialog management (DM), and user modeling (UM). Broadly speaking, ASR involves recognizing words from spoken audio for each user utterance, NLU is the process of determining the meaning of the recognized words in the utterance, DM involves generation of appropriate response utterances by the system and the management of the conversation with the user across multiple utterances, and UM is the process by which the system keeps track of information about a user, e.g., what their typical goals and preferences are.

41. These components are described in patents, papers, and textbooks published from the mid-1990s and 2000s, such as the well-known textbooks such as “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development” by Xuedong Huang, et al., published in 2001 (EX1012). For example, Huang provides a diagram of the architecture of a “spoken language understanding system,” containing ASR (“Speech Recognizer”), NLU (“Sentence Interpretation”), and DM (“Dialog Manager”). EX1012, Figure 1.4.

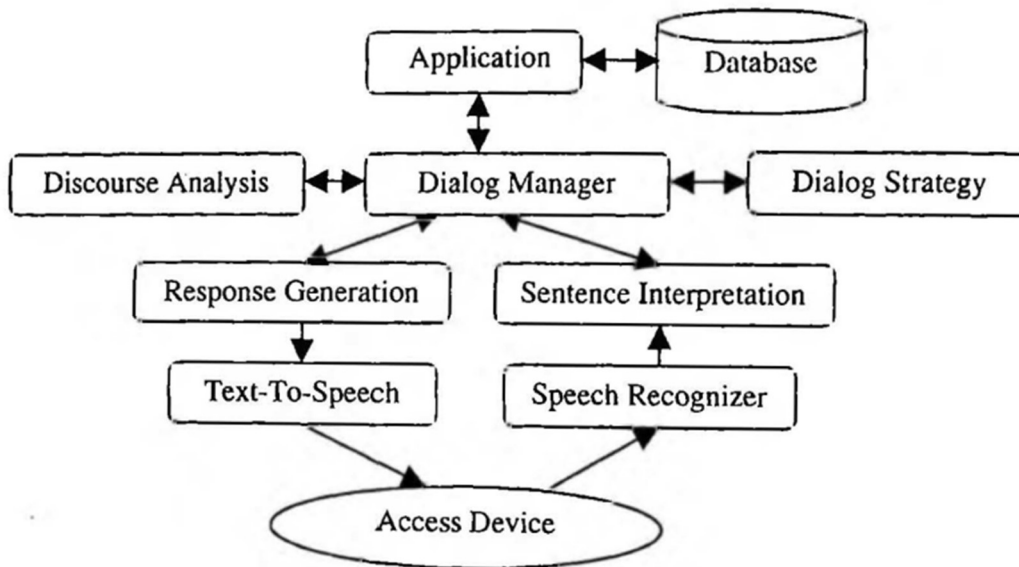


Figure 1.4 Basic system architecture of a spoken language understanding system.

42. Note that these terms are not the only terms that are used to describe these technologies. Other terms were (and are still) used by researchers and practitioners in this general context. For example, one such term is “Spoken Language Understanding” (SLU) which typically encompasses aspects of ASR, NLU, and DM. Nonetheless, the terms ASR, NLU, DM, and UM were in broad use in 2007 and would have been well-known to a POSITA at that time.

43. As of 2007, ASR typically involved a sequence of steps consisting of (i) recording and digitization of speech waveforms spoken by a user, (ii) conversion of the digitized speech waveforms into numerical features that (typically) represent the frequency content of very short segments of the speech waveform, (iii) automated classification of these features into corresponding phonemes (basic units of sounds within words), and (iv) recognition of specific words based on

sequences of phonemes. This process has been the subject of extensive research and commercial activity since the 1960s. For example, the Huang textbook devotes Chapters 5 through 13 (approximately 470 pages) just to the topic of ASR. EX1012, Chapters 5–13.

44. The output of the ASR component of a conversational interface system consists of the ASR's best estimates of what words were spoken by a user in an utterance. The recognized words alone are typically not sufficient for an algorithm to understand the user's goals and intent. Human language is very flexible and there are many ways to express the same meaning or intent. The technology of NLU (Natural Language Understanding, e.g., discussed in Chapter 17.3-5 in SLP) consists of a broad range of algorithms and methods that allow a computer system to interpret the meaning of the recognized words from the ASR. For example, consider the simple phrase "I would like to fly to New York" from the dialog example described earlier. And assume for example that the system is a travel reservation system that can find flights, make flight reservations, pay for the flights, suggest or advertise additional travel services, and so on. Such a system would have an internal representation (e.g., a database) of items such as names of airports, locations of airports, lists of flights, actions the system can take (e.g., book flights), and so on. Thus, it would need to interpret the user's words "like to fly" as an implicit request for the system to find a flight. It would also need to interpret the word "I" to mean

that the flight is for the individual who is speaking. And it would need to understand that “New York” is the name of a location, and that it is not the name of a specific airport but is a city with multiple airports associated with it. There are different approaches associated with developing NLU technologies that can robustly handle such situations, but by 2007 multiple such systems had been developed that demonstrated the necessary NLU capabilities for interpreting the meaning of spoken words in conversational dialog. *See, e.g.*, EX1012, Chapter 18; EX1011; EX1013, 18:39–19:37.

45. As described above, the ASR and NLU components of an interactive conversational system focus on recognizing words and interpreting these words on a per-utterance basis. DM (Dialog Management) technology addresses the problem of tracking different aspects of a dialog across multiple utterances. *See* EX1012, Chapter 17.6. For example, for a travel reservation system interacting with a user via speech, the DM component of the system needs to gradually accumulate multiple pieces of relevant information from the user (dates, times, arrival/departure airports, seating preferences, confirmation of booking, purchase information) over the course of a dialog, i.e., over a sequence of utterances generated the user. Thus, DM technologies involve creating internal representations of concepts such as goals for a user and methods to satisfy those goals. One such method would include directly asking the user for specific information, which in turn involves automatically

generating the appropriate words for the system's response (utterance) to the user, given what the system knows already (e.g., "What time would you like to leave for your flight to New York?") and then converting that into a spoken response that the user can hear. Another type of DM method could involve inferring the relevant information from knowledge sources available to the system. For example, the DM might determine where the user lives from information in the user's profile, and based on that information, suggest one or more departure airports within the geographic area of the user's home location. *See* EX1009, 20:4-15. And the DM could infer the likely destination airport if non-stop flights from the user's home airport only go to a specific airport in the New York area.

46. DM technology involves various software components that can make decisions about what type of information to request from the user, how to present this information to the user, in what sequence to request this information, when this information might be inferred from other sources, when sufficient information has been collected and/or inferred for the system to take a subsequent action, and so on. For example, in the example dialog in a travel reservation system above, the user may provide the needed information to search for a flight (e.g., departure city, destination city, and time) over multiple utterances, where the DM tracks which information has been provided and prompts the user for additional needed information after each utterance. In other cases, the DM may detect that the user has

provided multiple pieces of information in a single utterance, e.g., “I’d like to purchase a ticket to Cleveland this Friday at noon,” which combined with inferred information from the user’s file, e.g., the departure city being the user’s home city, is sufficient for the DM to initiate a search for available flights and proceed to a next stage of the dialog, e.g., presenting the user with available flight options and soliciting user confirmation to purchase. *See* EX1013, 18:49-60; EX1012, Chapter 17.6 (describing plan-based dialog management); EX1004, 9-10 (describing Dialogue Manager “employs a series of models that help it interpret the user input and decide on the next system action”). DM technology that exhibited these characteristics were well-known by 2007 had been developed in research laboratories and in commercial systems prior to that time. *See, e.g.,* EX1012, Chapters 17.6-17.7, Chapter 18; EX1011; EX1006; EX1004.

47. It was also well-known by 2007 that there were significant benefits to be gained if a conversational system had access to specific knowledge about a user (such as their travel preferences). Such access can in general make the interaction with the user both more efficient from the system’s perspective and a more pleasant experience from the user’s point of view, by reducing the amount of information that the system needs to request from the user during the dialog. The technology of user modeling (UM) consists of a number of different methods that can be used by a system to automatically build an internal representation of a user. These methods

can include relatively simple (but useful) approaches such as recording information about a user's full name, age, home address, etc., that is explicitly provided by a user during (say) a registration process. *See, e.g.*, EX1006, EX1009, EX1007, EX1004, EX1005. More complex UM methods extract information that is implicitly provided by the user, for example, inferences about a user's travel preferences from words in past conversations that a user has had with the system, or inferences from words in earlier utterances in the current conversation. *See, e.g.*, EX1006; EX1009; EX1004. As I discussed above, DM may make use of UM to infer information relating to a user's requests, without needing to request the information from the user. This may be used provide recommendations to the user and/or complete transactions on behalf of the user, e.g., using payment information previously provided to the user to complete a purchase transaction. *See, e.g.*, EX1007, EX1004, EX1009, EX1005. By 2007 the benefits of UM in conversational modeling were widely acknowledged and had been developed for various commercial systems. *See, e.g.*, EX1006, EX1007, EX1009, EX1004, EX1005.

48. In some speech systems, conversations between the user and the system take place within one or more domains or applications, such as stocks, news, music, travel planning, etc. *See, e.g.*, EX1012, Chapter 17; EX1008; EX1006; EX1004. Some systems provided access to multiple domains or applications, which a user could reach by specifically requesting a particular domain or application, or by

providing an utterance containing application- or domain-specific vocabulary that is mapped by the system to a particular application or domain, e.g., through keyword matching or other semantic analysis techniques. *See, e.g.*, EX1009, EX1006, EX1004. The different domains or applications may be associated with different types of tasks, objects, concepts, and behavior. For example, a travel reservation domain may require a more elaborate dialogue compared to, say, a domain for purchasing music CDs, as there is typically more information needed to be specified by the user before a search can be performed. *See* EX1009, 20:4-28. Thus, DM and UM may exhibit different behavior and perform different actions, depending on the domain or application of the conversation.

49. By 2007 it was also well-understood that the types of conversational systems that I described above, involving combinations of ASR, NLU, DM, and UM, could be directly applied to a variety of applications involving online commercial activities, such as advertising and purchasing of goods and services, in areas such as travel reservations, financial management, e-commerce purchases, and so on. *See, e.g.*, EX1012, Chapter 17.8; EX1011; EX1006; EX1007; EX1009; EX1013; EX1004 ; EX1005.

50. Thus by 2007, systems and methods for a conversational voice user interface were well-known. Indeed, as Kennewick recognized in 2004, several years before the earliest priority date of the challenged patent, “[m]uch work covering

multiple methods has been done in the fields of natural language processing and speech recognition. Speech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications. Natural language processing has previously been applied to the parsing of speech queries.” EX1006, [0006]. Speech recognition software was also commercially available. *See, e.g.*, EX1009, 9:55-59. Indeed, by 2001, long before the earliest priority date of the challenged patent, many different free voice portals were available that enabled users to access a variety of informational and e-commerce domains. *See, e.g.*, EX1012, Chapter 18, Table 18.1 (below). Thus, all the foundational technologies referenced in the claims were already known to POSITAs at the ’758 patent’s priority date.

Table 18.1 Some free voice portal features. These portals are being developed and will roll out more features.

Category	Audiopoint ⁵	Tell Me ⁶
Traffic	Yes	Yes
Weather	U.S. and world cities	U.S.
News	Yes	Yes
Financial	Yes	Yes
Sports	Yes	Yes
Airline info	No	Yes
Restaurants	No	U.S.
Entertainment	Yes	Yes
Personalization	Yes	Yes

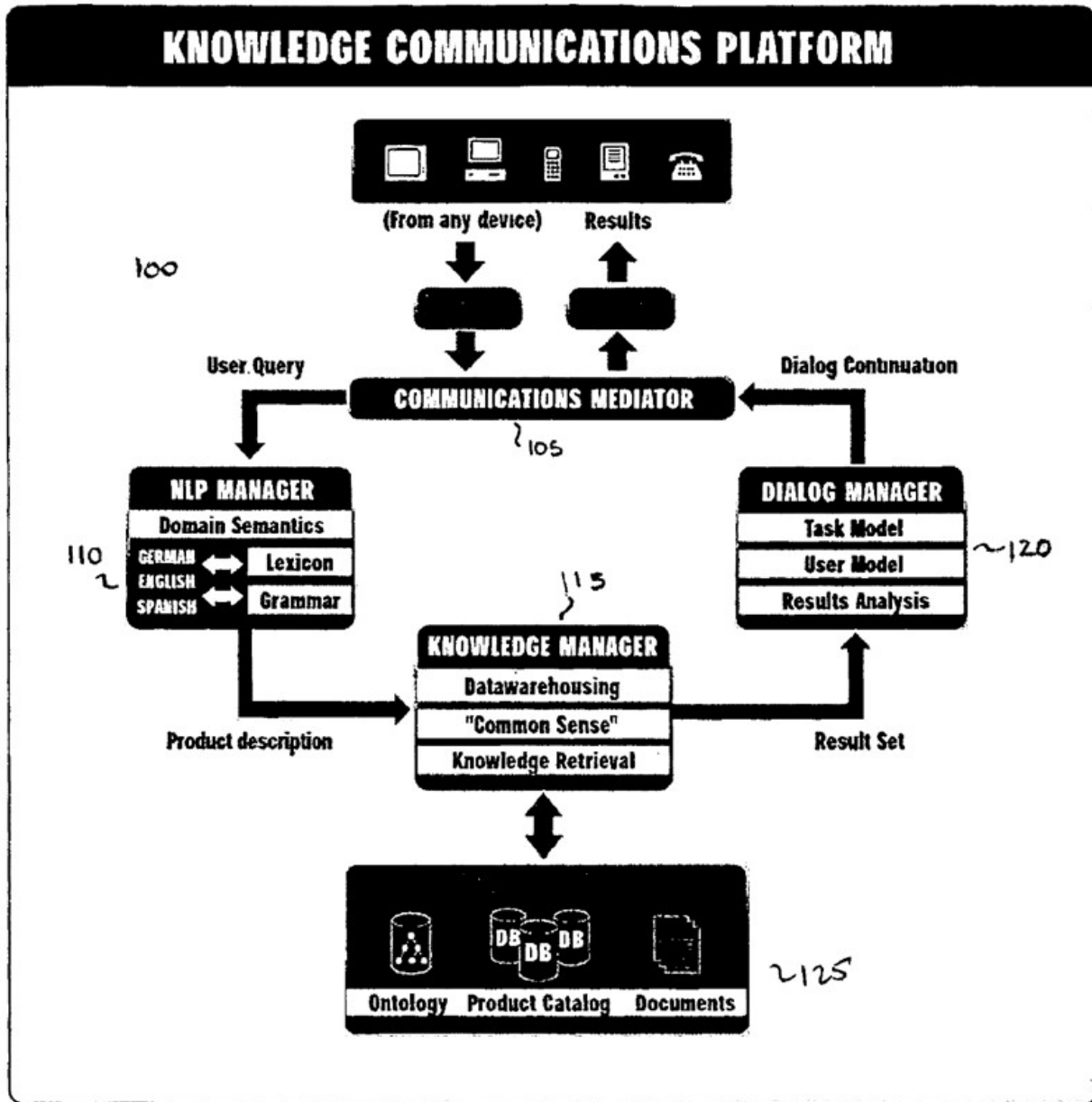
B. Summary of the Prior Art

1. Aretoulaki (EX1004)

51. Aretoulaki, WO Pub. No. 02/073331, is entitled “Natural Language Context-Sensitive and Knowledge-Based Interaction Environment for Dynamic And

Flexible Product, Service and Information Search and Presentation Applications.” Aretoulaki was filed on February 20, 2002, published on September 19, 2002, and claims earliest priority to February 20, 2001. EX1004. I understand Aretoulaki is prior art because its filing and publication dates predate the priority date of the challenged patent.

52. Aretoulaki discloses a “knowledge communications platform” for providing “intuitive natural language (text or speech) dialogue-based electronic commercial transactions and information exchanges.” EX1004, Abstract. FIG. 1 (reproduced below) illustrates that the “knowledge communications platform” includes a number of components called “processing modules,” including a “Communications Mediator, 105,” “NLP Manager, 110,” “Knowledge Manager, 115,” and “Dialogue Manager, 120.” EX1004, 7:50-8:8, FIG. 1.



53. Aretoulaki describes that the “Communications Mediator 105” receives user speech as user input 205 and translates the speech into an “input text string” (“ASCII Text Representation 215”) that it routes “to NLP Manager 110 for further processing.” EX1004, 8:9-18; *see also id.*, FIG. 2. The NLP Manager 110 analyzes the text representation to determine a domain or application associated with the

received utterance and acquire a “domain semantic representation 225” identifying “the desired product, service or information.” EX1004, 11:31-42. The Dialogue Manager 120 and Knowledge Manager 130 perform further “discourse” and “contextual analysis” to validate and disambiguate the “domain semantic representation” using prior discourse history and domain knowledge, to produce an updated “semantic representation 245” of the utterance. *Id.*, 12:15-52. For example, where “ambiguities” exist, e.g., if “the user employs a keyword that is relevant to more than a single domain or application . . . (e.g. a booking instruction or pricing information request),” the NLP Manager generates “a number of different interpretations . . . representing the various alternative meanings” of the utterance associated with the different possible domains. *Id.*, 15:1-29. The Dialogue Manager 120 and Knowledge Manager 125 consider the discourse history, user models, and “salient domain concepts in the knowledge base” to weigh the different interpretations and select one having “the highest weight.” *Id.*, 15:1-29.

54. Once the system establishes “the salient domain concepts and their values appearing in the user input,” the Knowledge Manager uses the resulting “product (or service) description” “for the dynamic database look-up” to retrieve a “Result Set” of products or services to present to the user. *Id.*, 8:27-46, 9:12-20, 13:37-40, 15:44-51. The user can then select one or more of the presented results to

“pose additional questions” about a product or service, and/or “ask to make a purchase (commit).” *Id.*, 22:20-23.

2. Ramer (EX1005)

55. Ramer, U.S. Patent No. 7,752,209, is entitled “Presenting Sponsored Content on a Mobile Communications Device.” Ramer was filed on January 19, 2006, issued on July 6, 2010, and claims earliest priority to September 14, 2005. EX1005. I understand Ramer is prior art because its filing date predates the priority date of the challenged patent.

56. Ramer discloses a “wireless search platform 100 for facilitating the access to and integration of multiple data sources and data types for presentation on a mobile communication facility 102,” an example of which is shown in FIG. 1 (below). EX1005, 3:64-67, 5:53-55.

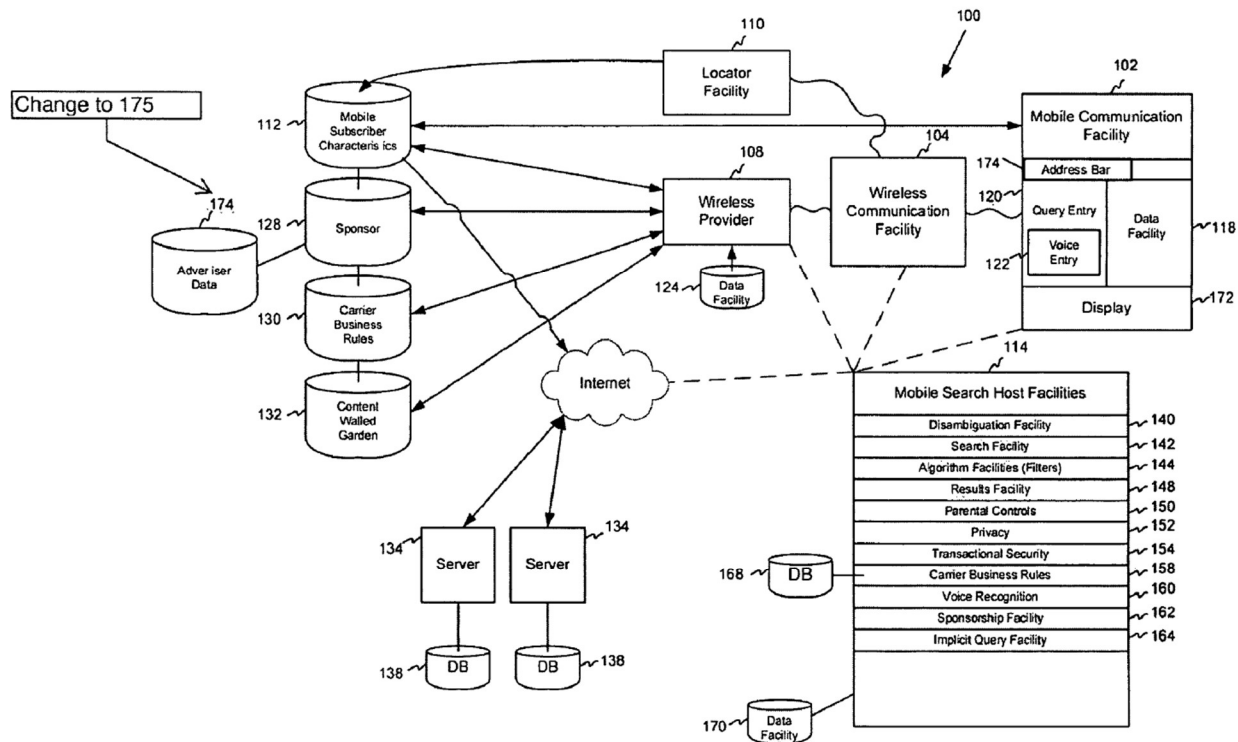


Fig. 1

57. The wireless search platform may receive a user’s “voice-based query entry 120.” EX1005, 53:3-6, 55:4-9; *see also* 15:47-49. The wireless search platform includes mobile search host facilities 114 that process and return results of received queries, including products available for purchase. The user can then select a product to purchase (e.g., “initiate . . . an online order), via clicking, keyboard, or “spoken commands” such as “buy it now.” *Id.*, 16:11-19, 21:25-28 (“voice recognition facility” enables system to “carry out spoken commands”); 42:21-24 (clicks, keystrokes, and “voice-initiated actions” are “user interface equivalents”), 47:63-48:13. For example, based on the received “query entry,” the system may

present the user with “a link” to “a commercial item that is available via a computer network,” e.g., “an item on an electronic commerce site” “associated with various e-commerce features, such as those enabling . . . single click purchasing” or “instant purchasing (e.g., ‘buy it now’).” *Id.*, 47:63-48:13. Ramer does not specify whether the user performs “instant purchasing” by click or by voice. To the extent Ramer discloses “instant purchasing” by click, a POSITA would have understood that it could also be performed by voice. That is because Ramer equates “voice-initiated actions” as “user interface equivalents” of “clicks and clickthroughs.” EX1005, 42:21-24. Thus, Ramer teaches that the “instant purchasing” can be performed by “spoken command” (*id.*, 21:25-28), like the user saying “buy it now.” *Id.* The user’s purchase may be delivered “through the wireless search platform,” which may provide “mobile wallet/billing-on-behalf-of” services for completing the transaction. *Id.*, 16:11-19, 124:54-56; *see also id.*, 8:52-56 (wireless provider facilitates “purchase of physical goods” by “collect[ing] the user’s credit card information” and “adding goods transactions automatically to the wireless provider’s bill to the user”).

58. Ramer’s wireless search platform includes a “disambiguation facility 140” configured to interpret and “deriv[e] greater clarity from ambiguous user queries” using information from a “mobile subscriber characteristics database 112.” EX1005, 23:60-24:7. The mobile subscriber characteristics database stores

“individual users’ voice commands and usage patterns,” “user preferences” such as “likes and dislikes” derived “through interpretation of other transactions . . . related to the mobile communication facility 102,” and/or other information, used to “assess the desired search targets of [the] user” and/or “refin[e] or customiz[e] the search to the particular user.” *Id.*, 6:21-27, 17:28-48, 17:61-8:4; *see also id.*, 24:9-16 (“the disambiguation facility 140 may receive the query from the wireless communication facility 104 or the wireless provider 108 and link the query to information known about the user that is stored in the mobile subscriber database 112 (e.g., age, sex, past Internet usage, etc.),” which, “coupled with the original query entry 120, may permit an unambiguous query to be processed”). The system may disambiguate queries and/or return query results “based on [the user’s] shopping habits.” *Id.*, 34:4-21. For example, the system may, responsive to a query for “King,” present “a frequent book purchaser . . . with Stephen King books,” “while a frequent music purchaser may be presented with results related to Elvis Presley.” *Id.*, 34:4-21; *see also id.*, 31:37-44 (“a user who has recently purchased a car and who enters ‘car’ in a user interface may receive information about registering the car, obtaining insurance, or the like, while a user who has recently shopped for, but not purchased, a car may receive results showing ratings of cars”).

3. Kennewick (EX1006)

59. Kennewick, U.S. Publication No. 2004/0193420, is entitled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance.” Kennewick was filed on July 15, 2003 and published on September 30, 2004. Kennewick claims earliest priority to July 15, 2002. EX1006. I understand Kennewick is prior art because its filing and publication dates predate the priority date of the challenged patent. I also understand that both Kennewick and the patent application that issued as the ’758 patent were initially assigned to VoiceBox Technologies Corp (“VoiceBox”).

60. Kennewick discloses a “speech interface” that enables “mobile users to ask natural language speech questions or give natural language commands in a wide range of domains.” EX1006, [0002]. The speech interface makes “significant use of context . . . and user specific profile data” to interpret a user’s request and provide responses. *Id.*, Abstract. As Figure 5 of Kennewick illustrates below, the user’s natural language utterance is received at a “main unit 98” that processes the utterance by a speech recognition engine 120 that “recognize[s] words and phrases.”

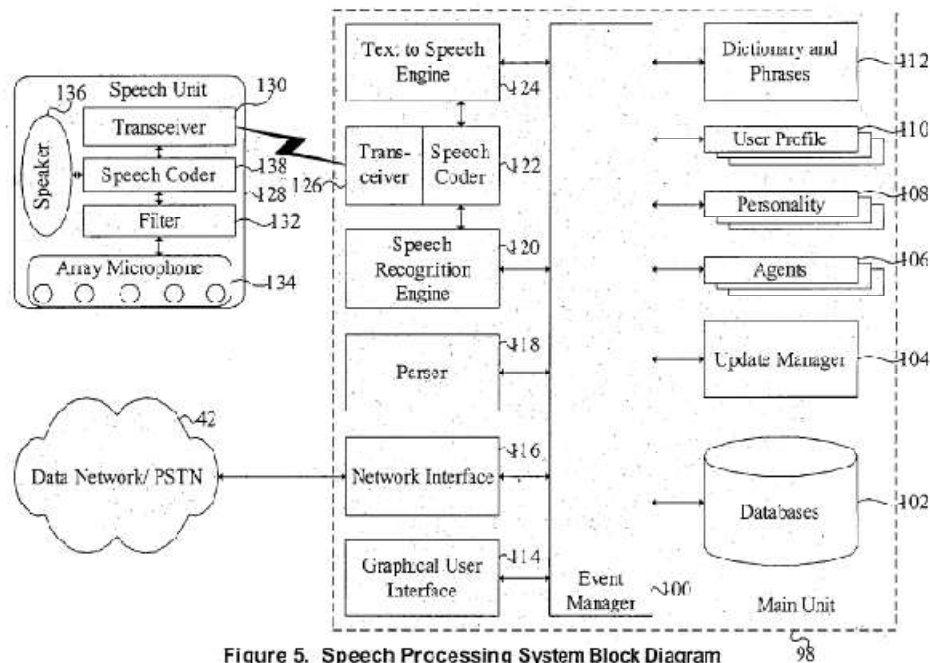


Figure 5. Speech Processing System Block Diagram

EX1006, FIG. 5, [0155].

61. The main unit further includes a parser 118 that determines a context for the user's utterance based on the recognized words and phrases of the utterance, by "applying prior probabilities or fuzzy possibilities to keyword matching." EX1006, [0160]. The parser 118 may further "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." *Id.*, [0160]-[0161]. For example, in one dialog "the keyword[] 'temperature' implies a context value of weather for the question" but in "a different dialog, the keyword 'temperature' can imply a context for a measurement" such as body temperature.

Id., [0160]. “Once the context for the question or command has been determined, the parser 118 can invoke the correct agent 156, 150.” *Id.*, [0162].

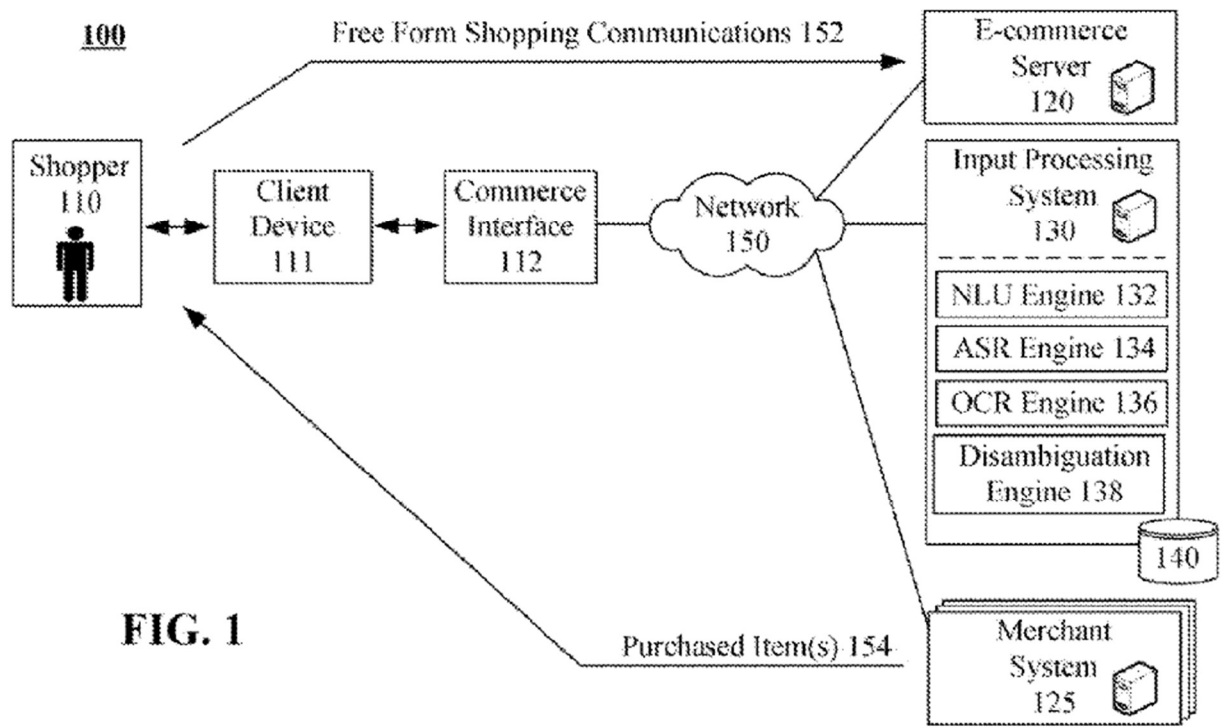
62. Additionally, Kennewick discloses that the speech interface may be used for various applications, including voice commerce applications. EX1006, [0060]-[0085]. For example, the speech interface may perform “[r]emote ordering and payment for goods and services,” in which the system provides “interactive offers and promotions for goods and services” and/or “[a] product catalog,” and facilitates “remote ordering by using location information, customer preferences, customer order history, etc.” and “a secure payment wallet for the users.” *Id.*, [0066]-[0067]. Kennewick further discloses different contexts or domains corresponding to different types of products, including “fast food ordering,” “[t]ravel services” and reservations, etc. *Id.*, [0018], [0068].

4. Hao (EX1007)

63. Hao, U.S. Publication No. 2008/0189187 is entitled “Populating an E-commerce Shopping Cart and Other E-commerce Fields Based Upon Content Extracted from Natural Language Input.” Hao was filed on February 1, 2007 and published on August 7, 2008. EX1007. I understand Hao is prior art because its filing date predates the priority date of the challenged patent.

64. Hao discloses solutions for “enhanc[ing] e-commerce shopping so that shoppers can interact with an automated system” by “submit[ting] request[s] . . . in

a natural language.” EX1007, [0009]. As shown in FIG. 1 (reproduced below), Hao’s input processing system includes “an automatic speech recognition (ASR) engine 134” and a “natural language understanding (NLU) engine 132” to process received speech input. *Id.*, [0022].



65. The NLU Engine extracts “semantic meanings” from the received user input, and a commerce server compares the extracted semantic meanings “to inventory items available to the shopper 110 for purchase from at least one merchant system 125.” EX1007, [0022]. The commerce server adds matching items “to an electronic shopping cart or other commerce artifact provided by the e-commerce server 120,” where the shopper can further “modify purchase details

and/or confirm a purchase via options provided by the commerce interface.” *Id.*, [0022].

66. Hao further discloses that when the user’s natural language input contains “ambiguities,” the system attempts to resolve the ambiguities “using shopper specific information maintained within a shopper profile/preferences data store 228.” EX1007, [0009], [0033]. For example, Hao’s FIG. 3 (reproduced below) illustrates a flowchart of “a method for conducting e-commerce.” *Id.*, FIG. 3.

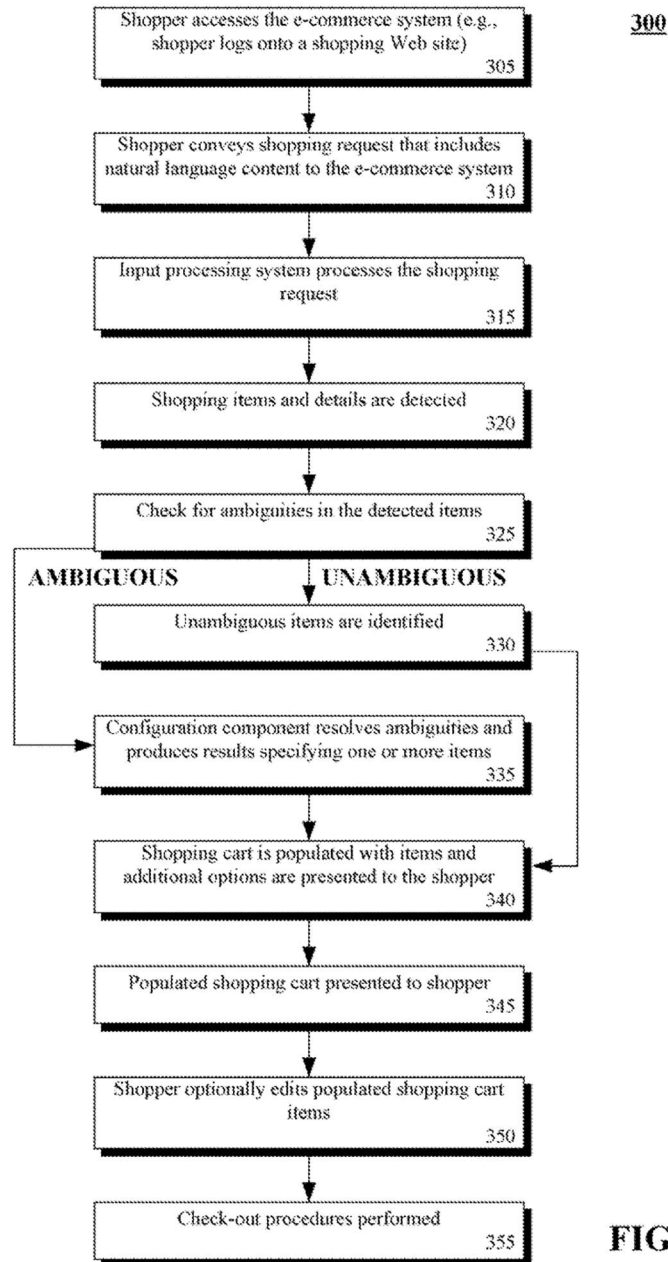


FIG. 3

67. As shown in FIG. 3, the system “check[s] for ambiguities” (325) and then attempts to resolve “ambiguous items” using “a configuration component” that utilizes “past transaction history . . . to increase accuracy of future transactions,” e.g., by “weigh[ing] items that have previously been purchased by a shopper when

resolving ambiguities.” EX1007, [0041]. For example, the system may resolve an ambiguous token of “cheese” in the user’s input based on the shopper’s “purchase history” which “indicate[s] that cheese refers to KRAFT 10 oz. packets.” *Id.*, [0046]. This automatic resolution of ambiguities streamlines the e-commerce process, by reducing a need to “prompt[] the shopper 110 to clarify their intended meaning” each time there is an ambiguity. *See id.*, [0028].

5. Jong (EX1008)

68. Jong, U.S. Patent No. 6,173,250, is entitled “Apparatus and Method for Speech-Text-Transmit Communication Over Data Networks.” Jong was filed on June 3, 1998 and issued on January 9, 2001. EX1008. I understand Jong is prior art because its filing and issue dates predate the priority date of the challenged patent.

69. Jong discloses a “speech recognition device[]” that translates “speech input signals . . . into text based on phonemes obtained from a spectral analysis of the speech input signals.” EX1008, Abstract. The speech recognition device uses “a listing of phonemes (speech sounds)” to identify “the phonemes in the speech input signal and extracts the corresponding syllables from [a] subword model database.” *Id.*, 5:51-67. The speech recognition device checks the syllables “against the lexicon data stored in [a] lexicon database” to group the syllables into “recognizable words.” *Id.*, 5:51-67.

VIII. SUMMARY OF INVALIDITY GROUNDS

70. In my opinion, the challenged claims are invalid based on the following grounds:

- Ground 1: Aretoulaki in combination with Ramer renders obvious claims 1-5, 9-10, 21-27, 31-32, and 43-44 of the '758 patent;
- Ground 2: Aretoulaki in combination with Ramer and Jong renders obvious claims 8 and 30 of the '758 patent;
- Ground 3: Aretoulaki in combination with Ramer and Kennewick renders obvious claims 6-7, 11-20, 28-29, and 33-42 of the '758 patent;
- Ground 4: Kennewick in combination with Hao renders obvious claims 1-2, 4-7, 9-24, 26-29, and 31-44 of the '758 patent;
- Ground 5: Kennewick in combination with Hao and Aretoulaki renders obvious claims 3 and 25 of the '758 patent; and
- Ground 6: Kennewick in combination with Hao and Jong renders obvious claims 8 and 30 of the '758 patent.

IX. INVALIDITY GROUNDS

A. Ground 1: Aretoulaki in combination with Ramer render obvious claims 1-5, 9-10, 21-27, 31-32, 43-44

1. General Motivations to Combine Aretoulaki and Ramer

71. In my opinion, a POSITA would have been motivated to combine Aretoulaki and Ramer. The references are in the same field of interactive speech

technology, and both relate to systems and methods for enabling users to conduct voice-based e-commerce to search for and purchase goods and services through an interactive speech interface. For example, Aretoulaki discloses a “system architecture and a methodology for intuitive natural language (text or speech) dialogue-based electronic commercial transactions and information exchanges.” EX1004, Abstract. The system analyzes received user queries using “Natural Language Processing (NLP),” to obtain a “product (or service) description which will be used for the dynamic database look-up,” the results of which are presented to the user, who can “pose additional questions . . . and/or ask to make a purchase (commit).” *Id.*, 8:27-48, 22:10-23; *see also id.*, 11:31-36 (user input “is processed lexically, syntactically, and semantically by NLP Manager 110 at block 220 in order to acquire a representation of the user's requirements as a domain semantic representation 225 (or some of these requirements) about the desired product, service or information”).

72. Ramer similarly discloses a “wireless search platform [] architecture” that “facilitates the processing of user-initiated [voice] queries,” including queries for purchasable products or services. EX1005, 4:7-10, 15:47-49. The system presents search results that “are targeted to the user based on information relating to the user,” and may include a “link to a commercial item that is available via a

computer network,” which the user can purchase through “single click purchasing” or “instant purchasing (e.g., ‘buy it now’).” *Id.*, 47:34-39, 47:63-48:13.

73. Additionally, both Aretoulaki and Ramer share similar architectures for an interactive speech system. Both disclose an input device that captures a user’s voice input. *See, e.g.*, EX1004, Abstract (user can “pose questions over the Internet” through “a PC microphone, a fixed or mobile phone,” etc.); EX1005, 6:21-29 (“The voice entry 122 function . . . may be used through the speaker-receiver device of the mobile communication facility 102”). Both disclose selecting and presenting a product or service to the user’s request based on the user’s voice input. *See, e.g.*, EX1004, 8:27-48 (processing “user queries” to determine a “product (or service) description which will be used for the dynamic database look-up”), 15:46-51 (presenting a “result set” of products or services “to the user . . . as either A[S]CII text, speech and/or images”); EX1005, 47:63-48:3 (“presenting a link in a user interface . . . to an item of goods or services for sale”). Both disclose enabling the user to purchase presented goods or services through the interactive speech system. *See, e.g.*, EX1004, 22:20-26 (“user can select” a product or service “proposed by the system and pose additional questions . . . and/or ask to make a purchase (commit)”); EX1005, 48:4-15 (presented link includes “various e-commerce features, such as those enabling single click purchasing, . . . instant purchasing (e.g., ‘buy it now’)”).

74. Both Aretoulaki and Ramer teach maintaining and updating user-specific information by tracking user interactions with the system, and using the maintained user-specific information to interpret received user inputs and select products or services for presentation to the user. For example, Aretoulaki teaches “employ[ing] a series of models” to help “interpret the user input and decide on the next system action.” EX1004, 9:22-42. These include the “discourse history” and a “user model contain[ing] data about the general preferences and assumptions of” the individual user. EX1004, 9:39-10:10. For example, the user model may include information such as a record of the user’s “previous orders,” which the system uses to “infer information and preferences about” the user when interpreting the user’s utterances, “speed[ing] up the interaction and the search process.” *Id.*, 14:4-18. Ramer also recognizes that received queries may be ambiguous and describes specific techniques to interpret and “deriv[e] greater clarity from ambiguous user queries” using information from a “mobile subscriber characteristics database,” such as the user’s “shopping habits” derived from their observed behavior. EX1005, 23:60-24:7, 34:4-21. Ramer provides further teachings of tracking different types of user interactions with presented products or services, including the user viewing goods and services and/or the completing a transaction related to a purchase opportunity. *See, e.g.*, EX1005, 8:36-52 (“shopping habits” derived from “views of or purchases of goods and services”), 12:21-24 (“[i]mplicit collection” of user data

“refers to data collected as . . . a byproduct of user behavior, such as products viewed . . . or products purchased”); 71:65-72:9 (items presented with “information to assist a user in making a purchase decision,” and with “an option to purchase these items”). Ramer uses this information to register user preferences and “shopping habits,” for use in disambiguating ambiguous user queries. EX1005, 23:60-24:7, 34:4-21.

75. A POSITA in possession of Aretoulaki and Ramer would therefore have been motivated to incorporate Ramer’s techniques for disambiguating ambiguous user queries based on user-specific information such as items purchased and shopping habits into Aretoulaki’s system, to enhance and improve the system’s ability to infer information about the user and “speed up the interaction and the search process.” For example, in the combined system, a POSITA would have found it obvious for Aretoulaki’s NLP Manager and Dialogue Manager, when interpreting a received natural language query, to consult a user-specific profile or model containing user preferences and shopping habits inferred from previously-tracked user interactions as taught by Ramer, to disambiguate the received query and select an appropriate purchase opportunity to provide to the user. *See, e.g.*, EX1004, 14:4-18; EX1005, 34:18-21. By resolving these ambiguities, the system can identify a result set of goods or services to present to the user as part of a selected purchase opportunity in a way that more closely matches “what users are looking for to satisfy their specific needs.” EX1004, 6:50-7:2. This modification is a combination of prior

art elements disclosed in Aretoulaki and Ramer according to known methods to yield predictable results. I provide additional motivations to combine the references below.

2. Claim 1

[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:”

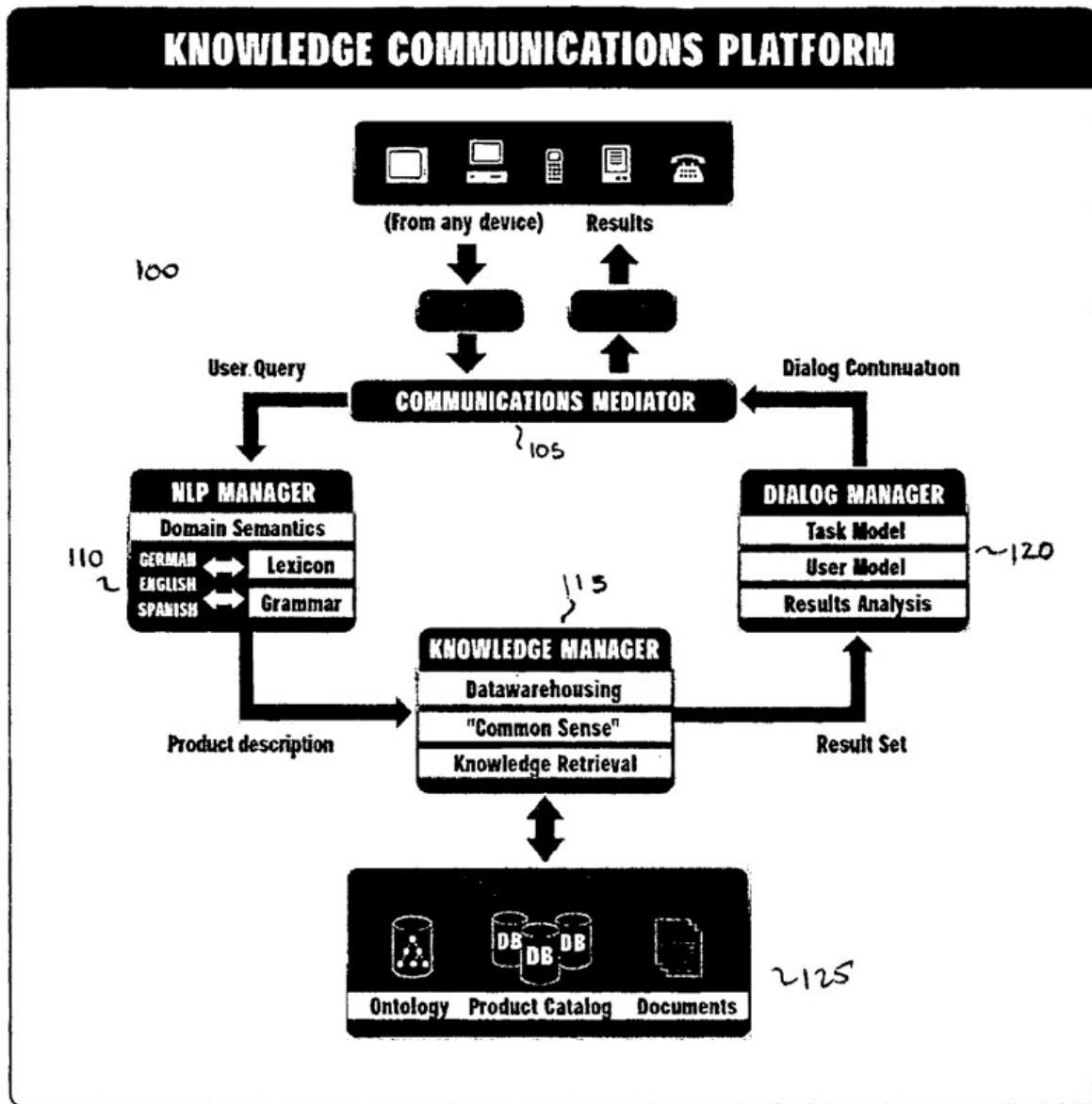
76. To the extent that the preamble of claim 1 is limiting, Aretoulaki in combination with Ramer renders it obvious. First, Aretoulaki discloses a “***method for processing natural language utterances that include requests.***” For example, Aretoulaki discloses a method “for intuitive natural language (text or speech) dialogue-based electronic commercial transactions and information exchanges.” EX1004, Abstract. The method includes processing “user queries” and/or “user requests” by interpreting them “in the context of the current interaction” to return an appropriate system response. *Id.*, Abstract, 8:27-46; *see also id.*, 4:9-11 (“system comprises modules for . . . the robust processing of natural language text or speech input from the user”), 6:18-21 (“use of . . . natural language input on the part of the user to interact with the system to find whatever he is looking for”). The processed “user queries”/“user requests” received during the natural language speech dialogue

as “natural language . . . speech” meet the claimed “*natural language utterances that include requests.*”

77. Second, Aretoulaki discloses “*selecting and presenting purchase opportunities based thereon*” (i.e., on the processed natural language utterances). For example, Aretoulaki describes processing the “user queries” (“*natural language utterances that include requests*”) to obtain a “product (or service) description” that the system uses to retrieve “a result set” of products or services to be presented to the user. *Id.*, 8:27-46, 9:12-20, 15:44-51. The selected set of products or services corresponds to the claimed “*purchase opportunities*” because the user is given the opportunity to “ask to make a purchase (commit)” of one or more of the presented products or services. *Id.*, 22:20-23. The system thus selects and presents the set of products or services (“*purchase opportunities*”) based on the user’s queries (i.e., “*natural language utterances that include requests*”).

78. Aretoulaki further renders obvious that the method is “*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.*” For example, Aretoulaki discloses that the method is implemented by a “knowledge communications platform,” which as shown in FIG. 1 (reproduced below), includes various “processing modules” such as a

“Communications Mediator,” “NLP Manager,” “Knowledge Manager,” and “Dialogue Manager.”



EX1004, 7:50-8:8, FIG. 1. The “knowledge communications platform” containing “processing modules” meets the claimed “*one or more physical processors*,” as a POSITA would have found it obvious for these processing modules to reside on some type of physical hardware capable of performing processing functions. In

addition, a POSITA would have found it obvious for the processing modules to operate based on computer code, e.g., “*computer program instructions*,” that when executed cause the processing modules to perform various method steps.

79. In addition, a POSITA would have found for the “knowledge communications platform” in the combination of Aretoulaki and Ramer to include “*one or more physical processors programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method.*” For example, Ramer discloses a “wireless search platform [] architecture” that, like Aretoulaki’s “knowledge communications platform,” “facilitates the processing of user-initiated [voice] queries.” EX1005, 4:7-10, 15:47-49. Ramer further discloses that the architecture may be deployed on “dynamically-scalable hardware and software to handle incoming requests,” such as clusters of web servers. *Id.*, 127:50-128:13. A POSITA would have understood that the clusters of web servers may correspond to “*one or more physical processors*,” because servers typically contain physical processors. In addition, Ramer discloses that the architecture implements various “software tool[s]” such as a “search facility” corresponding to “a high-volume, scalable, redundant search engine,” an “algorithm facility” for “evaluating a number of possible solutions based upon a user query,” a “disambiguation facility” for “complet[ing] or provid[ing] more meaning to ambiguous active user inputs,” etc. *Id.*, 9:8-19, 20:62-64, 38:42-44. A POSITA

would have understood that the software is implemented as computer code, e.g., “*computer program instructions*,” that when executed cause the hardware of the system to perform various method steps. In the combined system, a POSITA would have found it obvious that Aretoulaki’s “knowledge communications platform” containing various “processing modules” may be implemented on a “hardware and software” architecture similar to how Ramer’s “wireless search platform [] architecture” is implemented on “hardware and software.” EX1004, 7:50-8:8, FIG. 1; EX1005, 4:7-10, 15:47-49. For example, a POSITA would have found it obvious for the “processing modules” illustrated in FIG. 1 of Aretoulaki to be implemented on one or more servers. A POSITA would have found this modification to be obvious, as a combination of prior art elements according to known methods to yield predictable results, i.e., a system architecture implemented using one or more physical processors programmed with software. Aretoulaki in combination with Ramer therefore renders limitation [1.pre] obvious.

[1.a]: “providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;”

80. Aretoulaki in combination with Ramer renders this limitation obvious. As explained for limitation [1.pre] above, the “*one or more processors*” in Aretoulaki includes the “processing modules” of the “knowledge communications platform.” See EX1004, 7:50-8:7, FIG. 1. The processing modules include a “Communications Mediator 105” that handles “natural language input” received in

different modalities, including “speech,” and converts “the user input into text form.” *Id.*, 6:18-21, 8:9-16, 45:24-26. The natural language input received as speech meets the claimed “*natural language utterance*.” *See id.*, 6:18-21 (“use of . . . natural language input on the part of the user to interact with the system”). A POSITA would have also understood that the Communications Mediator includes the claimed “*speech recognition engine*,” because it receives input in the speech modality and translates the speech into recognized text. *See id.*, 5:31-37 (system “accepts any type of input that comes in - directly or indirectly - in a text form, as this is delivered by devices such as a computer keyboard, *speech recognizer*, etc.” emphasis added²), 8:16-18; *see also id.*, 10:36-39 (Communications Mediator translates “various input modes” into “an ASCII text representation 215 of user input 205”). Thus, Aretoulaki discloses that the Communications Mediator of the Knowledge Communications Platform receives a natural language speech utterance and “*provid[es]*” it to a “speech recognition engine” (e.g., “speech recognizer” implemented as part of the Communications Mediator) to convert the utterance “into text form.” Aretoulaki in combination with Ramer therefore renders limitation [1.a] obvious.

² All annotations and emphases are added unless otherwise noted.

[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;”

81. Aretoulaki in combination with Ramer renders this limitation obvious. Aretoulaki discloses that the “NLP Manager 110,” one of its processing modules (e.g., “*one or more processors*”), receives “*words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine.*” For example, as discussed above, Aretoulaki discloses that the Communications Mediator 105 includes the claimed “*speech recognition engine*” that converts a received natural language utterance into recognized text, e.g., an “ASCII text representation 215 of user input 205.” EX1004, 5:31-37, 8:9-18, 10:36-39. This recognized text corresponds to the claimed “*words or phrases, recognized from the natural language utterance, as an output of the speech recognition engine.*” The Communications Mediator “routes” the “input text string” corresponding to the recognized text “to the NLP Manager 110 for further processing.” *Id.*, 8:9-18; *see also id.*, 8:42-43 (NLP Manager processes the “recognized words from the user input”), 29:9-14 (discussing NLP components of dialogue system mapping words or phrases of the user input to a domain ontology). Thus, the NLP Manager of Aretoulaki’s system (e.g., “knowledge communications platform” having “*one or more processors*”) receives the words or phrases recognized from the natural language utterance as an output of the speech recognition engine. Aretoulaki in combination with Ramer therefore renders limitation [1.b] obvious.

[1.c] “determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;”

82. Aretoulaki in combination with Ramer renders this limitation obvious. Aretoulaki discloses that the system’s processing modules (e.g., ***“one or more processors”***), including the NLP Manager 110, process the “text representation 215 of the user input 205” corresponding to the received natural language utterance and ***“determin[e] a context for the natural language utterance based on the recognized words or phrases”*** of the text representation. See EX1004, 11:31-36. For example, the NLP Manager analyzes the recognized words or phrases to acquire a “domain semantic representation 225” of the natural language utterance indicating “the user’s . . . desired product, service or information,” by identifying “individual words” that “point to domain specifications.” *Id.*, 11:31-42. The NLP Manager utilizes “a mapping lexicon that translates application-specific words into the corresponding relevant concepts in the domain ontology.” *Id.*, 11:42-45. Thus, Aretoulaki discloses determining ***“a context for the natural language utterance”*** by determining an application or domain associated with the utterance, by mapping the text (***“recognized words or phrases”***) of the utterance to a specific domain specification and ontology for generating a “domain semantic representation” of the utterance. See also *id.*, 12:15-19 (describing further “discourse analysis” performed on “initial interpretation of the user input” to yield a “domain and dialogue act semantic representation”).

83. Aretoulaki's disclosure of determining a context associated with the utterance by determining a domain is consistent with the '758 patent's disclosure of determining a context associated with the utterance. For example, Aretoulaki discloses determining the context for the utterance may include generating "a number of different interpretations . . . representing the various alternative meanings in the case of ambiguities," e.g., if "the user employs a keyword that is relevant to more than a single domain or application . . . (e.g. a booking instruction or pricing information request)," and selecting one having "the highest weight." EX1004, 15:16-29. This is consistent with the disclosure of the '758 patent, which describes determining "a context for a current utterance 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. In other words, like Aretoulaki, the '758 patent indicates that a context for a natural language utterance may correspond to an application or domain associated with the utterance. *See also* EX1004, 15:16-29 (describing different domains such as "booking" and "pricing"); EX1001, 9:36-56 (describing that the context of an utterance may correspond to an application or domain such as "Music," "Navigation," or "Search"). Aretoulaki in combination with Ramer therefore renders limitation [1.c] obvious.

[1.d] “selecting, by the one or more processors, a purchase opportunity based on the determined context;”

84. Aretoulaki in combination with Ramer renders this limitation obvious. Aretoulaki discloses that the system’s processing modules (e.g., “***one or more processors***”) process the text representation of the received natural language utterance to select “***a purchase opportunity based on the determined context.***” For example, the NLP Manager 110 processes the utterance text based on the determined context/domain “to acquire . . . a domain semantic representation 225” representing the user’s requirements for “the desired product, service or information.” EX1004, 11:31-40. The Dialogue Manager 120 may further analyze and augment the “domain semantic representation” to generate a “final Domain Semantic Representation 255” establishing “the salient domain concepts and their values appearing in the user input.” *Id.*, 12:15-52, 15:1-50. Based on this representation, The Knowledge Manager 115 “dynamically accesses the application database 125” and “attempts to retrieve the best matches to the user’s requirements,” and the Dialogue Manager presents the results to the user in the form of “text, speech and/or images.” *Id.*, 15:44-51. The presented result set corresponding to the best matching products or services meets the claimed selected “***purchase opportunity***” because the user is given the opportunity “make a purchase (commit)” of one or more of the presented products or services. *See id.*, 22:10-26. In other words, Aretoulaki discloses determining the context, e.g., a domain or application, of the user input, to

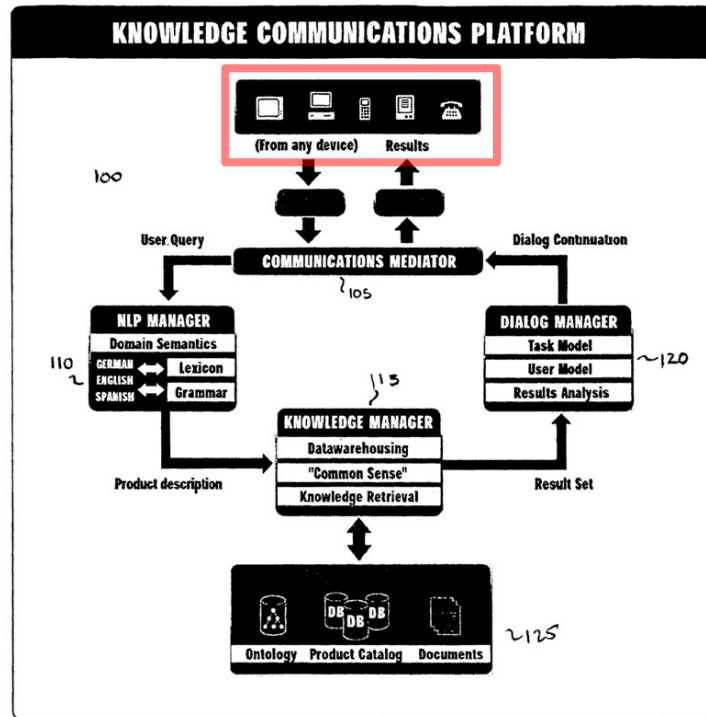
form a “domain semantic representation” of the user input, and uses the generated representation to access one or more databases and retrieve a best match to the user’s request based thereon. Thus, Aretoulaki discloses selecting “*a purchase opportunity based on the determined context*,” because it discloses selecting the purchase opportunity by searching an application database using the domain semantic representation generated based on the determined context/domain. Aretoulaki in combination with Ramer therefore renders limitation [1.d] obvious.

[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;”

85. Aretoulaki in combination with Ramer renders this limitation obvious. First, Aretoulaki discloses that the system’s processing modules (e.g., “*one or more processors*”) deliver “*the selected purchase opportunity*” to the user. As discussed above, Aretoulaki discloses that after the system establishes “the salient domain concepts and their values appearing in the user input,” the Knowledge Manager “accesses the application database 125 . . . and attempts to retrieve the best matches to the user’s requirements as specified up to that point.” EX1004, 15:1-9, 15:44-51. The system delivers the resulting matches corresponding to a “retrieved product, service and[/or] other information” to the user as “as either A[S]CII text, speech and/or images.” *Id.*, 15:44-51, 25:9-14 (describing “multimodal presentation of the retrieved product, service and other information that best matches the user requirements ensures that the user will be interested in finding out more about it or

even make a related purchase”). The delivered results meet the claimed “*selected purchase opportunity*” because the user can interact with “and/or ask to make a purchase (commit)” of the retrieved product or service. *Id.*, 22:20-23.

86. Aretoulaki further discloses delivering the selected purchase opportunity “*via an electronic device in communication with the one or more processors.*” For example, Aretoulaki discloses that the Communications Mediator “coordinates the presentation of” the system’s output, including the delivered purchase opportunity, 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; *see also id.*, 25:9-14 (describing “multimodal presentation of the retrieved product, service and other information”). Thus, Aretoulaki discloses delivering the selected purchase opportunity to the user via “*an electronic device*” such as a phone and/or an electronic device having a screen, or a personal computer or PDA, that is in communication with the Communications Mediator 105 (of the “*one or more processors*”). *See also id.*, 4:4-7, 8:12-14 (user interacts with the system through a “PC, mobile or normal phone, or devices such as a PDA, etc.”). For example, FIG. 1 (reproduced below) illustrates various electronic devices in communication with the Communications Manager via which the system may deliver the purchase opportunity (“Results”). Aretoulaki in combination with Ramer therefore renders limitation [1.e] obvious.



[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;”*

87. Aretoulaki in combination with Ramer renders this limitation obvious.

First, Aretoulaki discloses *“tracking, by the one or more processors, an interaction pattern associated with the purchase opportunity delivered to the electronic device.”* For example, Aretoulaki discloses tracking “all the exchanges in a specific interaction session with the user” to maintain and update a “user model” containing “data about the general preferences and assumptions of” the individual user. EX1004, 9:47-10:5, 22:28-34 (“evolving” user model based on data “collected in

real time”). A POSITA thus would have understood that Aretoulaki discloses “*tracking . . . an interaction pattern*” because Aretoulaki discloses tracking (e.g., monitoring) all interactions between the user and the system to build the user model, including any interactions with presented purchase opportunities.

88. Aretoulaki further discloses that the interaction pattern is “*associated with the purchase opportunity delivered to the electronic device.*” For example, Aretoulaki discloses that the monitored user behavior may include a record of the user’s “previous orders,” which the Dialogue Manager may use to “interpret the user input and decide on the next system action.” *Id.*, 9:39-49, 14:4-8. In other words, Aretoulaki discloses tracking (e.g., monitoring) “*an interaction pattern associated with the [delivered] purchase opportunity,*” e.g., to track which products or services (e.g., “database matches proposed by the system”) associated with delivered purchase opportunities the user has purchased, to record the user’s “previous orders.” *Id.*, 22:20-26 (describing that user may interact with a delivered purchase opportunity by “select[ing] one . . . of the [proposed] database matches” and “ask[ing] to make a purchase (commit)” of the product/service). Furthermore, because Aretoulaki discloses tracking (e.g., monitoring) all interactions between the user and the system to build the user model and discourse history, a POSITA would have found it obvious that the tracked interactions include tracked interaction patterns associated with any purchase opportunity delivered during a conversation.

89. Aretoulaki further discloses that the tracked interaction pattern ***“includes information associated with a subsequent request in which the electronic device interacts with the purchase opportunity.”*** As discussed above, the user, upon reviewing a delivered purchase opportunity (e.g., “database matches” of products or services “proposed by the system”) via the electronic device (e.g., the user’s PC, telephone, or PDA), may make one or more subsequent requests through the electronic device such as “select[ing] one . . . of the database matches proposed by the system,” “pos[ing] additional questions about” the presented products or services, “and/or ask[ing] to make a purchase (commit).” EX1004, 22:20-26; *see also id.*, 8:12-16 (describing user inputs received via “PC, mobile or normal phone, or devices such as a PDA”), 25:9-14 (“user will be interested in finding out more about [the retrieved product or service] or even make a related purchase by getting directed to the corresponding E-commerce site”). For example, the user may make a request “to buy or book a product or service just presented by the system” as part of a delivered purchase opportunity, by making a natural language request such as “Great! I’d like to book that,” “Sounds good. How can I pay?,” or “Excellent. Will I get a receipt directly sent to my home address?” *Id.*, 40:22-35. This request meets the claimed ***“subsequent request in which the electronic device interacts with the purchase opportunity,”*** since the user’s interaction with the delivered purchase opportunity occurs through the electronic device. *See id.*, 8:12-16 (user interacts

with the system through a “PC, mobile or normal phone, or devices such as a PDA, etc.”).

90. Aretoulaki further discloses that “*the subsequent request . . . includes the electronic device completing a transaction related to the purchase opportunity.*” As discussed above, the claimed “subsequent request” may include the user, through the electronic device, making a request “to make a purchase (commit)” of a product or service associated with the delivered purchase opportunity. Aretoulaki further discloses that when this request is made, the system enables the user to, through the electronic device, “*complet[e] a transaction related to the purchase opportunity*” to make the purchase, such as by redirecting the user to a “corresponding e-commerce site.” *Id.*, 22:20-26, 25:9-14. Thus, Aretoulaki discloses tracking (e.g., monitoring) an interaction pattern associated with a delivered purchase opportunity that includes information of a subsequent request in which the electronic device interacts with the purchase opportunity, e.g., the user request to purchase the selected product or service (commit), and includes the electronic device completing a transaction related to the purchase opportunity, which is reflected in the record of the user’s “previous orders” included in the user model. *See id.*, 14:4-8.

91. In addition, to the extent that Aretoulaki does not explicitly disclose “*tracking . . . an interaction pattern associated with the [delivered] purchase*

opportunity” that includes information associated with a subsequent request “*completing a transaction related to the purchase opportunity*,” the combination with Ramer renders it obvious. Ramer discloses a “wireless search platform” that “facilitat[es] the access to and integration of multiple data sources and data types for presentation on a mobile communication facility” corresponding to an “*electronic device*” such as a “cellular phone, Blackberry, wireless electronic mail device, [and/or] personal digital assistant” that the user uses to interact with the wireless search platform. EX1005, 3:64-67, 5:10-25, 5:53-55. For example, a user may submit “a voice-based query entry 120” through the “mobile communication facility” for processing by the wireless search platform. *Id.*, 53:3-6, 55:4-9; *see also* 15:47-49 (“query entry” can be made via voice or text). In response, the wireless search platform returns query results corresponding to products available for purchase, e.g., “a link . . . to an item of goods or services for sale” “associated with various e-commerce features, such as . . . instant purchasing (e.g., ‘buy it now’).” *Id.*, 47:63-48:13. The presented products correspond to a “*purchase opportunity*,” where the user may interact with the purchase opportunity by “clicking on a sponsored link or engaging in a commercial transaction, such as purchasing [] content.” *Id.*, 53:14-20, 71:65-72:9 (describing presenting items with “information to assist a user in making a purchase decision,” and with “an option to purchase these items”); *see also id.*, 8:52-56 (wireless provider facilitates “purchase of physical

goods” by “collect[ing] the user’s credit card information” and “adding goods transactions automatically to the wireless provider’s bill to the user”); 16:11-19 (user after viewing results may “initiate . . . an online order” delivered “through the wireless search platform”).

92. Ramer discloses that the system tracks “*an interaction pattern associated with the [delivered] purchase opportunity*” by tracking user activities “such as purchasing a product online, visiting a site on line, . . . viewing content, or engaging or not engaging in other activities” to draw inferences on the user’s preferences. EX1005, 10:41-50 (“if the user looks at a [presented] product and decided not to purchase,” the system may “draw an inference that the user is not interested in the product”). For example, the delivered purchase opportunity presented to a user may include a link “to an item on an electronic commerce site,” where the link is associated “with various e-commerce features, such as those enabling single click purchasing, . . . instant purchasing (e.g., ‘buy it now’), tracking of clicks or transactions, tracking of referrals, affiliate program features, or the like.” *Id.*, 48:4-13. Ramer thus discloses tracking the “on-line interactions of [the] mobile communications facility” (i.e., the “*electronic device*”) with the delivered purchased opportunity, by “tracking clicks, clickthroughs, queries, . . . purchases, payment timing,” etc. *Id.*, 124:5-21. 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 makes a purchase of the item associated with the purchase opportunity.

93. In the combined system, to the extent not explicitly disclosed by Aretoulaki, a POSITA would have found it obvious to modify Aretoulaki to track user activities such as completing a transaction related to a purchase opportunity as taught by Ramer, in order to register user preferences and/or to build a user model that tracks user purchase history. As discussed above, the references are analogous, both relating to systems and methods for enabling users to conduct voice-based e-commerce to search for and purchase goods and services through an interactive speech interface. *See, e.g.*, EX1004, Abstract, 8:27-48, 11:31-36, 22:10-23; EX1005, 4:7-10, 15:47-49, 47:34-39, 47:63-48:13. The combination is a combination of prior art elements disclosed in Aretoulaki and Ramer according to known methods to yield predictable results. For example, Aretoulaki teaches “employ[ing] a series of models” to help “interpret the user input and decide on the next system action,” including a “user model contain[ing] data about the general preferences and assumptions of” the individual user EX1004, 9:22-10:10. The user model includes information of the user’s “previous orders,” which Aretoulaki’s system uses to interpret subsequent user inputs by “infer[ring] information and preferences about” the user. *Id.*, 14:4-18. This “is useful in the setting of defaults

or the clarification of ambiguities, which speed up the interaction and the search process.” *Id.* Ramer in turn explicitly teaches tracking different types of interactions, including the completion of a transaction related to the purchase opportunity. *See, e.g.*, EX1005, 12:21-24 (“[i]mplicit collection” of user data “refers to data collected as . . . a byproduct of user behavior, such as products viewed . . . or products purchased”); 71:65-72:9 (items presented with “information to assist a user in making a purchase decision,” and with “an option to purchase these items”). Ramer uses the tracked information to register user preferences and “shopping habits” for use in “deriving greater clarity from ambiguous user requests.” EX1005, 23:60-24:7, 34:4-21. A POSITA in possession of Aretoulaki and Ramer would therefore have been motivated to incorporate Ramer’s techniques of tracking different types of interactions associated with a delivered purchase opportunity and registering user preferences and shopping habits based on those interactions, to enhance and improve Aretoulaki’s ability to infer information about the user and “speed up the interaction and the search process.” 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 knowledge communications platform tracking different types of interactions associated with a delivered purchase opportunity as taught in Ramer with which to “infer information and preferences about” the user. EX1004, 14:12-15. A POSITA would also have found the

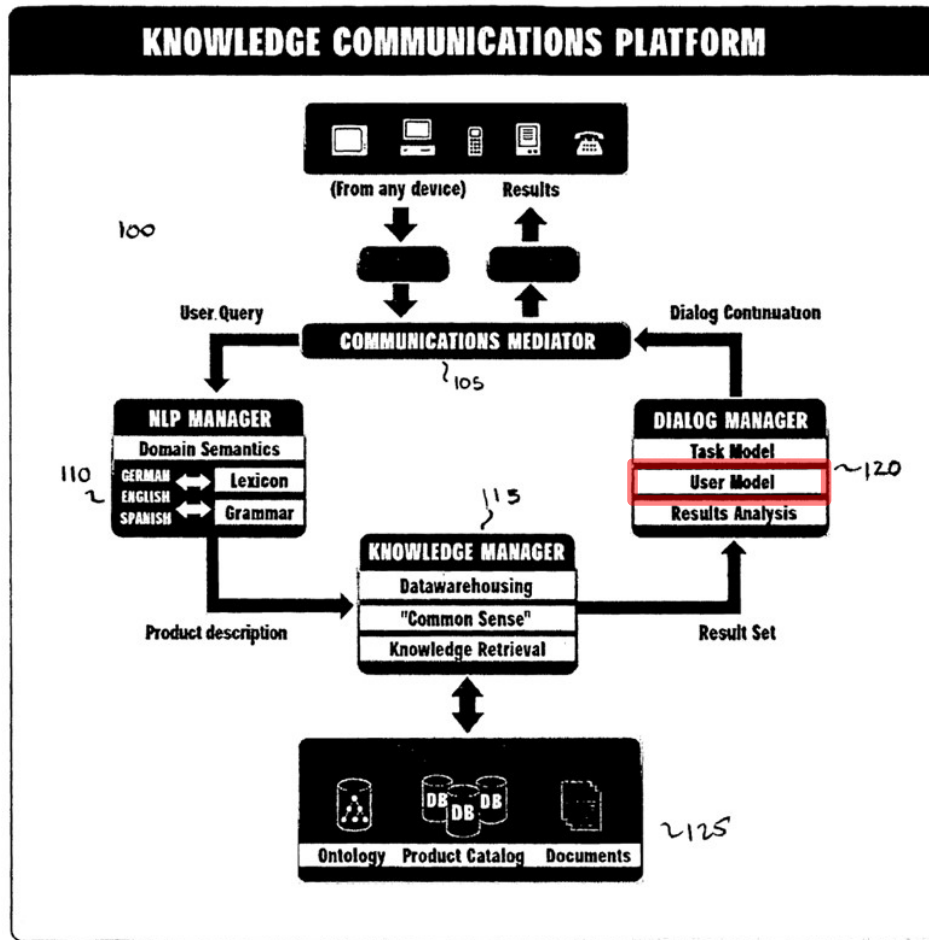
modification obvious to try as a way to create a more robust and natural interface.

The combination of Aretoulaki and Ramer thus renders limitation [1.f] obvious.

[1.g] “building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and”

94. Aretoulaki in combination with Ramer renders this limitation obvious.

First, Aretoulaki discloses “***building or updating . . . a user-specific profile based on the tracked interaction pattern.***” For example, Aretoulaki discloses that the system (e.g., “***the one or more processors***”), which includes the “Dialogue Manager 120,” builds and updates a user model based on the tracked interaction pattern. *See* EX1004, 9:39-10:10. For example, the user model “contains data about the general preferences and assumptions of” the individual user that may evolve over time based the user’s behavior “that is monitored in the course of the interaction,” including the user’s “previous orders.” EX1004, 9:47-10:5, 14:1-8, 22:28-34. FIG. 1, reproduced below, illustrates the “user model” as being maintained by the dialogue manager.



95. A POSITA would have understood that Aretoulaki’s user model correspond to a “*user-specific profile*,” as it contains “preferences and assumptions” about the specific user derived from the user’s observed behavior. *See id.*, 9:47-10:10 (user model may be “evolving” based on “monitored” user behavior “collected in real time”). A POSITA would also have found it obvious that the system builds or updates the user-specific profile “*based on the tracked interaction pattern*,” because the user-specific profile includes a record of the user’s “previous orders,” and a POSITA would have found obvious for the system to update the user’s

“previous orders” whenever the user makes a new order by interacting with a presented purchase opportunity to purchase a product or service.

96. In addition, to the extent that Aretoulaki does not explicitly disclose ***“building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern,”*** the combination with Ramer renders it obvious. Ramer discloses ***“building or updating . . . a user-specific profile”*** corresponding to “mobile subscriber characteristics 112 pertaining to individual users,” which includes data for a specific user, such as “user transaction history,” “content presented for viewed by/not viewed by user,” and “[u]ser preferences . . . derived from user behavior or other implicit characteristics, or explicitly defined by [the user].” EX1005, 2:45-64, 8:25-56, 10:24-50. Ramer discloses building or updating the profile by registering user preferences obtained by tracking interaction patterns such as “clickthroughs, queries,” “products purchased,” or the user “engaging or not engaging in other activities.” EX1005, 10:41-50 (“[u]sers may implicitly register a preference through activity”), 12:14-26, 124:13-21. For example, FIG. 3 of Ramer illustrates a mobile subscriber characteristics database 112 (storing the claimed ***“user-specific profile”***) that is built or updated using data received from the wireless provider 108, and is accessed by the disambiguation facility 140 to disambiguate received queries. *See id.*, 23:60-24:28

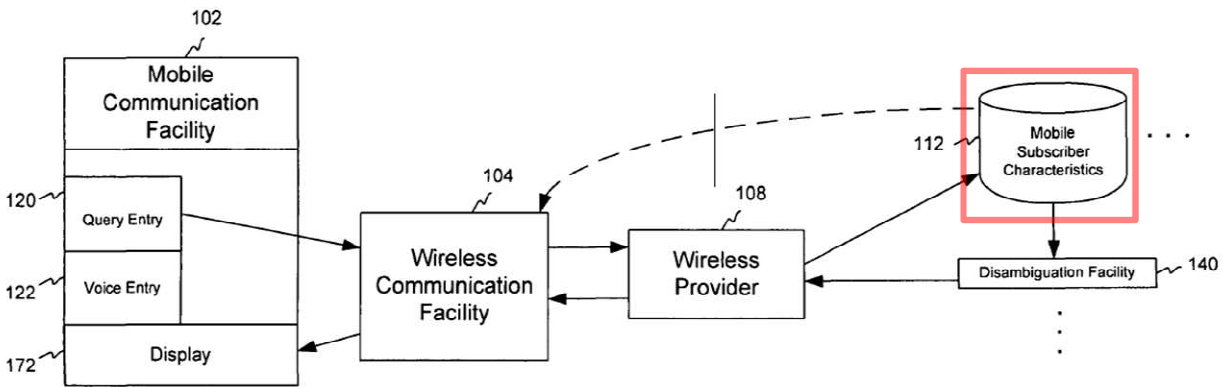


Fig. 3

97. In the combined system, a POSITA would have found it obvious to modify Areoulaki’s “user model,” to the extent that it does not already, to be updated for each user based on the tracked user interaction patterns associated with delivered purchase opportunities. For example, a POSITA would have understood that the combined system may derive the user’s preference information maintained in the user model “from [tracked] user behavior or other implicit characteristics” including interaction patterns where the user “purchas[es] a product online” as taught by Ramer. *See* EX1005, 10:34-50; *see also* EX1004, 9:47-10:2 (“user model contains data about the general preferences and assumptions of individual users,” based on user behavior “monitored by the system in the course of [an] interaction”). The system builds and updates the user’s derived preferences as the user interacts with presented purchase opportunities as taught by Ramer and uses the information in the user model to disambiguate user queries and “make predictions regarding

other information/products that might interest the user.” EX1005, 12:14-16. A POSITA in possession of Aretoulaki and Ramer would have been motivated to combine the references in this way, to enhance and improve Aretoulaki’s ability to make “assumptions” about the user to “speed up the interaction and the search process” and “avoid[] tedious questioning of the user.” See EX1004, 14:12-18. A POSITA would have found this modification obvious as a combination of prior art elements according to known methods to yield predictable results, and would have found the modification obvious to try as a way to create a more robust and natural interface. The combination of Aretoulaki and Ramer thus renders limitation [1.g] obvious.

[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.”

98. Aretoulaki in combination with Ramer renders this limitation obvious. First, Aretoulaki discloses “*interpreting, by the one or more processors, a subsequent natural language utterance.*” For example, Aretoulaki discloses that the NLP Manager and the Dialogue Manager of the system (e.g., the “*one or more processors*”) interpret received natural language utterances “in terms of . . . the context of the interaction, i.e. the previous exchanges between system and user.” EX1004, 12:43-49. Thus, a POSITA would have found it obvious that the system interprets a “*subsequent*” natural language utterance, e.g., based on previous

exchanges between the system and user including prior natural language utterances. *See also id.*, 23:12-14 (user may “choose their preferred expression medium” to interact with the system, such as “speech”).

99. Aretoulaki further renders obvious interpreting the subsequent natural language utterance “*using the user-specific profile.*” For example, Aretoulaki discloses that the Dialogue Manager of the system (e.g., the “*one or more processors*”) employs the “user model contain[ing] data about the general preferences and assumptions of” the individual user, which meet the claimed “*user-specific profile,*” to “help it interpret the user input and decide on the next system action.” EX1004, 9:39-10:10. Aretoulaki uses this profile, which contains data of the user’s history of “previous orders,” to interpret subsequent user inputs by “infer[ring] information and preferences about” the user, which “is useful in the setting of defaults or the clarification of ambiguities.” *Id.*, 14:4-18. This “speed[s] up the interaction and the search process,” and “avoid[s] tedious questioning of the user.” *Id.*; *see also id.*, 12:43-52 (“previous exchanges” and “data on the specific user” are used to perform “contextual analysis” to yield a “validated and disambiguated” interpretation), 25:33-36 (interaction “is personalized, because of the user models and personal profiles maintained and constantly updated by the system with each new session”).

100. In addition, a POSITA would have understood that Aretoulaki may interpret the subsequent natural language utterance using the user-specific profile to *“select a subsequent purchase opportunity based on the subsequent natural language utterance.”* For example, as discussed above, Aretoulaki discloses that the Dialogue Manager uses the user model, if available, to “interpret” each received utterance “and decide on the next system action.” EX1004, 9:22-52. Based on the interpretation, the Knowledge Manager 115 “dynamically accesses the application database 125 . . . and attempts to retrieve the best matches to the user’s requirements as specified up to that point,” and presents the result to the user as a purchase opportunity. *Id.*, 15:44-51. Because the system selects the purchase opportunity based on the user’s utterance interpreted using the user model containing the user’s “previous orders” relating to previously delivered purchase opportunities, a POSITA would have understood that the selected purchase opportunity is a *“subsequent purchase opportunity.”* Thus, a POSITA would have found it obvious that the system interprets the subsequent natural language utterance using the user-specific profile to *“select a subsequent purchase opportunity based on the subsequent natural language utterance.”*

101. In addition, to the extent that Aretoulaki does not explicitly disclose interpreting a subsequent natural language utterance using the user-specific profile *“to select a subsequent purchase opportunity based on the subsequent natural*

language utterance,” the combination with Ramer renders it obvious. For example, Ramer discloses using “information from a user’s profile to make predictions regarding other information/products that might interest the user,” where the user profile information includes previous “products purchased” by the user. EX1005, 12:14-24. Ramer further discloses using the user profile information to interpret a received user voice query (e.g., a “*subsequent natural language utterance*”) to select a purchase opportunity based on the query, by using the “*user-specific profile*” (i.e., “mobile subscriber characteristics”) during a disambiguation process to “deriv[e] greater clarity from ambiguous user queries.” EX1005, 23:60-66. For example, if the user provides “a query entry 120 ‘Coltrain,’” and the user’s profiles shows “a history of many recent purchases of compact discs recorded by the saxophonist, John Coltrane,” the system may take this “user-specific data contained in the [user’s] purchase history” into account to interpret the user’s query as seeking “information pertaining to ‘Coltrane,’ as opposed to ‘Coal Train’ or ‘Soul Train.’” EX1005, 24:65-25:17.

102. By interpreting the user’s query corresponding to the claimed “*subsequent natural language utterance*” using the user-specific profile, Ramer’s system selects a “*subsequent purchase opportunity*” based on the interpreted query. For example, a user whose profile indicates that they are “a frequent book purchaser may be presented with Stephen King books” in response to a query for “King,”

whereas a user who is “a frequent music purchaser may be presented with results related to Elvis Presley.” *Id.*, 34:18-21; *see also id.*, 31:39-44 (“a user who has recently purchased a car and who enters ‘car’ in a user interface may receive information about registering the car . . . while a user who has recently shopped for, but not purchased a car may receive results showing ratings of cars”). In other words, Ramer discloses interpreting the user’s subsequent input using a user-specific profile built or updated through tracking the user’s past interactions (e.g., past purchases) to select a subsequent purchase opportunity based thereon (e.g., interpreting “King” to select books by “Stephen King” if the user is a frequent book buyer). *See also id.*, 24:65-25:17 (interpreting user utterance as relating to “John Coltrane,” based on user’s “history of many recent purchases”).

103. In the proposed combination, a POSITA would have found it obvious to modify Aretoulaki to utilize user-specific profile information including past purchase information as taught in Ramer to interpret received user queries (including a “***subsequent natural language utterance***”) and select a purchase opportunity (e.g., a “***subsequent purchase opportunity***”) based thereon. For example, Aretoulaki discloses using the user model to “infer information and preferences” about the user and clarify “ambiguities” present in the user’s utterance. EX1004, 14:4-18; *see also id.*, 12:43-52 (“previous exchanges” and “data on the specific user” are used to yield a “validated and disambiguated” interpretation). Ramer also teaches using user-

specific profile information to interpret user inputs as part of a disambiguation process to “deriv[e] greater clarity from ambiguous user queries.” EX1005, 23:60-66. Ramer further provides specific examples of ways to interpret user inputs using user-specific profile information, such as the user’s determined “shopping habits” and “items purchased” history, to infer the desired meaning of the user’s input and select appropriate purchase opportunities in response. *See id.*, 23:60-66, 31:39-44, 34:18-21. For example, in the combined system, a POSITA would have found it obvious for Aretoulaki’s NLP Manager and Dialogue Manager, when interpreting a “subsequent natural language utterance” that includes the term “King,” to disambiguate the user’s intended meaning by accessing shopping habit data from the user-specific profile as taught by Ramer, and select an appropriate “subsequent purchase opportunity” (e.g., books by Stephen King if the user is a frequent book purchaser), consistent with Aretoulaki’s teaching of using the user model to “infer . . . preferences about” the user. EX1004, 14:4-18; EX1005, 34:18-21.

104. A POSITA in possession of Aretoulaki and Ramer would have been motivated to combine the references in this way, to enhance and improve Aretoulaki’s ability to “speed up the interaction and the search process” by using user-specific profile information to automatically disambiguate ambiguous user inputs, reducing a need for additional dialog turns to ask the user for clarification. A POSITA would further have been motivated to make the modification as it aligns

with Aretoulaki's teachings of using user model information "to infer information and preferences about" the user, which is "useful in the setting of defaults or the clarification of ambiguities, which speed up the interaction and the search process, avoiding tedious questioning of the user." EX1004, 14:12-18. A POSITA would have had a reasonable expectation of success in making this modification, as it involves a combination of prior art elements (e.g., Aretoulaki's teaching of using user profile information to interpret user requests, and Ramer's more specific examples of disambiguating user requests using user profile information) according to known methods to yield predictable results (e.g., Aretoulaki's system interpreting user requests using the specific disambiguation techniques taught by Ramer), and would have found the modification obvious to try as a way to create a more robust and natural interface. The combination of Aretoulaki and Ramer thus renders limitation [1.h] obvious.

3. Claim 2

[2.pre] "The method of claim 1, wherein the purchase opportunity relates to an item, the method further comprising:"

105. Aretoulaki in combination with Ramer renders the preamble obvious. For example, Aretoulaki discloses that the delivered purchase opportunity may relate to a product that is an item, such as a printer, scanner, skiing equipment, etc. *See, e.g.,* EX1004, 6:52-7:2, 11:1-10, 13:18-35; *see also* EX1005, 47:63-48:13 (describing purchase opportunities for "an item of goods or services for sale").

[2.a] “receiving, by the one or more processors, a second natural language utterance;”

106. Aretoulaki in combination with Ramer renders this limitation obvious. For example, as discussed for limitation [1.a], Aretoulaki discloses that the processing modules of the system (e.g., “**one or more processors**”) include a “Communications Mediator 105” that receives user speech inputs meeting the claimed “**natural language utterance.**” EX1004, 8:9-16, 6:18-21 (“use of . . . natural language input on the part of the user to interact with the system”). Aretoulaki further discloses that the Communications Mediator may receive multiple natural language utterances. *See, e.g., id.*, 8:9-16 (“It receives user queries . . .”). For example, when the system presents the user with a purchase opportunity, the user can “pose additional questions about” an item associated with the purchase opportunity, “and/or ask to make a purchase (commit).” *Id.*, 22:20-23. The user’s request “to buy or book a product or service just presented by the system” may be in the form of a natural language request such as “Great! I’d like to book that,” “Sounds good. How can I pay?,” or “Excellent. Will I get a receipt directly sent to my home address?” *Id.*, 40:22-35; *see also id.*, 4:11-12 (user interacts with the system through “natural language text or speech input”). Because this utterance is received after the delivery of the purchase opportunity selected based on the “natural language utterance” of claim 1, this natural language utterance is “**a second natural language**

utterance.” The combination of Aretoulaki and Ramer therefore renders limitation [2.a] obvious.

[2.b] “determining, by the one or more processors, that the second natural language utterance comprises a request to purchase the item; and”

107. Aretoulaki in combination with Ramer renders this limitation obvious. As discussed above, Aretoulaki discloses that when the system presents a purchase opportunity to the user, the user can “ask to make a purchase (commit),” which meets the claimed “**second natural language utterance.**” EX1004, 22:20-23. Aretoulaki discloses that the processing modules of the system (e.g., “**one or more processors**”) include an NLP Manager 110 and Dialogue Manager 120 that process each received natural language utterance to decide “on an interpretation of the user intention in saying” the utterance. *Id.*, 8:9-16, 12:30-35. The system expresses the determined user intention as a “user dialogue act[]” or “Dact,” where a user’s request to make a purchase corresponds to a “Commit” Dact indicating “that the user is committed to buy or book a product or service just presented by the system.” *Id.*, 23:21-23, 31:4-5, 40:22-24. Thus, Aretoulaki discloses “**determining, by the one or more processors**” (e.g., the NLP Manager and Dialogue Manager) “**that the second natural language utterance comprises a request to purchase the item,**” by determining that the utterance corresponds to a “Commit” Dact. The combination of Aretoulaki and Ramer therefore renders limitation [2.b] obvious.

[2.c] “completing, by the one or more processors, a purchase transaction of the item based on the second natural language utterance.”

108. Aretoulaki in combination with Ramer renders this limitation obvious. As discussed above, Aretoulaki discloses determining that the second natural utterance comprises a request to purchase the item. Aretoulaki further discloses that upon determining the user’s request to purchase the item, Aretoulaki’s knowledge communications platform (e.g., the “***one or more processors***”) enables the user to complete the purchase by redirecting the user “to the corresponding E-commerce site . . . through a URL link.” EX1004, 25:9-14. Thus, a POSITA would have found it obvious for the system to complete the purchase transaction of the item “***based on the second natural language utterance***” because Aretoulaki discloses completing the purchase using functionality that is “smoothly integrated in the invented architecture,” enabling the system to conduct “dialogue-based electronic commercial transactions,” responsive to determining that the second natural language utterance is a request to purchase. *Id.*, Abstract, 22:20-26. To achieve its stated purpose of enabling “dialogue-based *electronic commercial transactions*,” it would be necessary for the system to complete a purchase transaction. *Id.*

109. To the extent that Aretoulaki does not explicitly disclose “***completing . . . a purchase transaction of the item***,” the combination with Ramer renders it obvious. Ramer discloses that when the system presents results corresponding to selected products or services to the user, “the user may initiate

interactions/transactions 234 with the information,” such as “an online order” to complete the purchase. EX1005, 16:15-19 (purchase may be delivered “wirelessly through the wireless platform”). The user can select a product to purchase, via clicking, keyboard, or voice commands (“*second natural language utterance*”) such as “buy it now.” *Id.*, 48:8-13, 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”). For example, Ramer discloses providing the user with a link “to an item on an electronic commerce site,” where the user completes the purchase using “various e-commerce features,” such as “single click purchasing, [or] instant purchasing (e.g., ‘buy it now’).” *Id.*, 48:4-13. A POSITA would have found it obvious for the user’s request to “buy it now” to be received as a “spoken command[]” (“*second natural language utterance*”) because Ramer discloses that the user may provide “spoken commands” to the system through a “voice recognition facility,” and recognizes clicks, keystrokes, and “voice-initiated actions” as equivalent ways for the user to interact with the system. *Id.*, 21:25-28, 42:21-24. A POSITA would further have understood that the system performing instant purchasing based on the utterance “buy it now” corresponds to “*completing . . . a purchase transaction of the item based on the second natural language utterance.*”

110. In response to the user's command, the system may facilitate the "purchase of physical goods . . . 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." *Id.*, 8:52-56. The system may provide "mobile wallet/billing-on-behalf-of" services to complete the transaction, enabling the user to purchase "content through [the] wireless provider." *Id.*, 124:54-56, 136:33-39. Thus, Ramer discloses the system completing a purchase and billing a user for the purchase, responsive to the user initiating a purchase transaction based on the delivered purchase opportunity.

111. In the proposed combination, a POSITA would have found it obvious to modify Aretoulaki's knowledge communications platform, to the extent Aretoulaki does not explicitly disclose it, to, responsive to receiving the user's "***second natural language utterance***" requesting to "purchase (commit)," complete a purchase transaction for the item associated with the purchase opportunity as taught by Ramer. For example, a POSITA would have found it obvious to modify Aretoulaki's system such that, responsive to determining that the user's "***second natural language utterance***" (e.g., "buy it now") indicates a "Commit" intention, it then completes the purchase, e.g., using e-commerce features such as instant purchasing and/or by accessing a "mobile wallet" to collect user credit card information with which to bill the user as taught by Ramer. *See* EX1005, 8:52-56,

16:15-19, 48:6-13, 124:54-56. A POSITA would have been motivated to make this modification to enable Aretoulaki's system, to the extent not explicitly disclosed, to complete a purchase transaction for the item based on the user's utterance indicating their intention to purchase, enabling Aretoulaki's system to achieve its stated purpose of providing "intuitive natural language (text or speech) dialogue-based *electronic commercial transactions*." EX1004, Abstract. A POSITA would have found the combination obvious as a simple substitution of one known element (e.g., completing the purchase as taught by Ramer using instant purchasing and/or by billing the user) for another to obtain predictable results, as 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. EX1005, 8:52-56, 16:15-19, 21:25-28, 48:8-13, 124:54-56. The combination Aretoulaki and Ramer therefore renders limitation [2.b] obvious.

4. Claim 3

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

112. Aretoulaki in combination with Ramer renders the method of claim 1 obvious.

[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;”

113. Aretoulaki in combination with Ramer renders this limitation obvious. As discussed above, Aretoulaki discloses tracking “***user interactions with [a] purchase opportunity***,” where the user can “correct the system, or to interrupt it in the case of a misunderstanding,” in order to “locate communication or understanding problems or modified wishes.” EX1004, 7:11-19, 7:37-41. The system monitors the user’s interactions to detect “[b]reaks in the normal flow, such as . . . objections on the part of the user or erroneous recognition on the part of the system,” and in response “trigger a repair mechanism” to correct the misunderstanding. *Id.*, 18:19-25; *see also id.*, 32:21-25 (monitoring user interactions to identify “a correction of [a] parameter value”).

114. 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” or tracking the user’s interactions with the system, including interactions with any presented purchase opportunities. *See* EX1004, 7:10-24. Aretoulaki’s Communications Mediator (of the “***one or more processors***”) receives “***an indication of***” the occurrence via the electronic device the user uses to interact with Aretoulaki’s knowledge communications platform. *See id.*, 8:9-13 (Communications Mediator

“handles input and output of different modalities”). The combination Aretoulaki and Ramer therefore renders limitation [3.a] obvious.

[3.b] “determining, by the one or more processors, whether the interpretation of the natural language utterance was correct based on the indication; and”

115. Aretoulaki in combination with Ramer renders this limitation obvious. As I discussed above, Aretoulaki discloses that the Dialogue Manager (of the “*one or more processors*”) determines “*whether the interpretation of the natural language utterance was correct based an indication*” of user interaction with the purchase opportunity, because it monitors user interactions with the purchase opportunity to identify any “[b]reaks in the normal flow” indicating “objections on the part of the user or erroneous recognition on the part of the system.” EX1004, 18:19-25. For example, for a received “indication” of a user interaction, Aretoulaki’s “Dialogue Manager 120 decides on an interpretation of the user intention” in making the interaction, expressed as a “user dialogue act[,]” which may be a “Correct” dialog act 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 claimed “*natural language utterance*”). *Id.*, 12:30-35, 38:44-51. Thus, Aretoulaki’s system determines “*whether the interpretation of the natural language utterance was correct based on the indication,*” by determining whether the indication of the user’s interaction corresponds to a “correct” dialog act. The combination Aretoulaki and Ramer therefore renders limitation [3.a] obvious.

[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.”

116. Aretoulaki in combination with Ramer renders this limitation obvious. As discussed above, Aretoulaki discloses that the Dialogue Manager (of the “***one or more processors***”) triggers a “repair mechanism” responsive to determining, based on the indication, that the interpretation of the utterance was incorrect, e.g., by determining that the user’s interaction with the purchase opportunity corresponds to a “Correct” dialog act indicating the user intends to “correct the instantiation for a parameter” due to it being “misrecognized or misinterpreted.” EX1004, 18:19-25, 38:44-51. The repair mechanism may include the system updating its interpretation of the natural language utterance, e.g., establishing a “new/modified parameter value.” *Id.*, 15:37-42, 39:1-5.

117. In addition, Aretoulaki discloses “***interpreting . . . a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct,***” because it interprets each utterance “in terms of the preceding exchanges between the user and the system.” *Id.*, 4:17-23; *see also id.*, 16:38-40 (“The generated expectations about the next user input and its meaning in the context of what has taken place before guide the interpretation of the [current] input”), 22:28-31 (“The system always keeps a record of all the exchanges in a specific interaction session with the user (discourse

history) . . . to contextually interpret each new user input”). For example, following execution of the repair strategy to correct an incorrectly interpreted parameter, Aretoulaki would interpret subsequent utterances in the context of the “new/modified parameter value.” *See id.*, 39:1-5. Thus, the system interprets the “*subsequent natural language utterance*” based on the determination of whether the interpretation of the (prior) natural language utterance was correct, by interpreting it in the context of the corrected parameter. This is consistent with the disclosure of the ’758 patent, which explains “previous utterances” determined to have been interpreted incorrectly “may be reinterpreted to improve subsequent interpretations”. EX1001, 5:28-32. The combination Aretoulaki and Ramer therefore renders limitation [3.c] obvious.

5. Claim 4: “The method of claim 1, wherein delivering the selected purchase opportunity via the electronic device comprises: audibly presenting, by the electronic device, a natural language conversational purchase opportunity.”

118. Aretoulaki in combination with Ramer renders claim 4 obvious. First, as discussed for limitation [1.e], Aretoulaki discloses delivering the selected purchase opportunity via the electronic device. For example, Aretoulaki discloses that after analyzing the natural language utterance, the knowledge manager “accesses the application database 125 . . . and attempts to retrieve the best matches to the user’s requirements as specified up to that point.” EX1004, 15:1-9, 15:44-51. The system delivers the matches, which correspond to the “*selected purchase*

opportunity,” to the user in the form of “A[S]CII text, speech and/or images.” *Id.*, 15:44-51. Thus Aretoulaki discloses delivering the selected purchase opportunity by “*audibly presenting*” it as a “*natural language conversational purchase opportunity*,” because the purchase opportunity may be delivered as “speech.” The delivered “speech” may, in the form of natural language conversational speech, enable the system to provide a “natural language dialogue-based electronic commercial transaction.” *Id.*, 7:50-52. The combination Aretoulaki and Ramer therefore renders claim 4 obvious.

6. Claim 5

[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:”

119. Aretoulaki in combination with Ramer renders the preamble obvious. Aretoulaki’s system includes processing modules such as an “NLP Manager,” “Dialogue Manager,” and “Knowledge Manager.” *See* EX1004, 7:50-8:7, FIG. 1. As I discussed above, a POSITA would have understood that these processing modules operate based on computer code, e.g., “*computer program instructions*,” which configures them to perform various method steps. These modules correspond to the claimed “*conversational language processor*” that interprets a natural language utterance. *Id.*, 8:4-7. For example, the NLP Manager processes an “input text string” of a received natural language utterance, and outputs “frames with

embedded structures holding the recognized words from the user input, their syntactic and semantic function and possible semantic relationships among them.” *Id.*, 8:9-16, 8:40-44. These frames are an initial interpretation of the natural language utterance. *See also id.*, 8:29-35 (“Natural Language Processing (NLP) involves the lexical, syntactic (grammatical), and domain semantic analysis of the user input using both statistical observations of the various surface forms and a deeper interpretation of the relationships and dependencies among words, phrases, and concepts.”). The Dialogue Manager may further interpret this initial interpretation by augmenting the initial interpretation through discourse analysis, e.g., using “user dialogue acts.” *Id.*, 12:15-35; *see also id.*, 9:39-44 (“Dialogue Manager 120 employs a series of models that help it interpret the user input and decide on the next system action.”), 15:7-9 (“final Domain Semantic Representation 255”). The Dialogue Manager and Knowledge Manager may also select between “different interpretations . . . representing the various alternative meanings in the case of ambiguities,” using “[i]nformation about the context of the exchange . . . kept in the discourse history, and about the salient domain concepts in the knowledge base.” *Id.*, 15:16-30. Thus, Aretoulaki discloses an “NLP Manager,” “Dialogue Manager,” and “Knowledge Manager” that form a “*conversational language processor*” that is “*configured to interpret a natural language utterance*,” as these components work together to interpret the natural language utterance and establish an

interpretation of the utterance containing “the salient domain concepts and their values appearing in the user input,” which the system uses for selecting a purchase opportunity. *Id.*, 15:44-51. The combination Aretoulaki and Ramer therefore renders limitation [5.pre] obvious.

[5.a] “providing, by the one or more processors, the words or phrases as an input to the conversational language processor; and”

120. Aretoulaki in combination with Ramer renders this limitation obvious. For example, Aretoulaki discloses that the processing modules of the system (e.g., the “*one or more processors*”) include a “Communications Mediator” that translates the natural language utterance into “an ASCII string” corresponding to the recognized “*words or phrases*” of the utterance, and provides it to the NLP Manager “for further processing.” EX1004, 8:9-18, 10:36-39. As discussed above, Aretoulaki’s NLP Manager, Dialogue Manager, and Knowledge Manager form the claimed conversational language processor. *See also id.*, 8:40-44, 9:22-31. Thus, Aretoulaki discloses “*providing, by the one or more processors, the words or phrases as an input to the conversational language processor,*” because it discloses the Communications Mediator (of the “*one or more processors*”) providing the recognized text of the natural language utterance to the NLP Manager, which is part of the claimed “*conversational language processor.*” *See also id.*, 11:31-33 (“The ASCII text representation 215 of the user input 205 is processed lexically,

syntactically, and semantically by NLP Manager 110”). The combination Aretoulaki and Ramer therefore renders limitation [5.a] obvious.

[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.”

121. Aretoulaki in combination with Ramer renders this limitation obvious. As discussed above, Aretoulaki discloses that the system’s processing modules (e.g., the “*one or more processors*”) include an “NLP Manager,” “Dialogue Manager,” and “Knowledge Manager,” which form the claimed “*conversational language processor*.” Aretoulaki further discloses “*receiving . . . an interpretation of the natural language utterance, based on the words or phrases, as an output of the conversational language processor*.” As discussed above, Aretoulaki’s NLP Manager receives the words or phrases recognized from the natural language utterance as an “input text string,” and processes it to output “frames with embedded structures holding the recognized words from the user input, their syntactic and semantic function and possible semantic relationships among them.” EX1004, 8:12-18, 8:9-16, 8:40-44. The frames “make[] up the so-called product (or service) description which will be used for the dynamic database look-up,” and may be further processed by the Dialogue Manager and Knowledge Manager, to locate and resolve any “inconsistencies, incompatibilities, and contradictions in the evolving specification of the user's requirements (the updated product (or service)

description).” *Id.*, 8:44-46, 9:6-12. The generated frames thus correspond to the received “*interpretation of the natural language utterance*,” as they specify the “product (or service) description” reflecting the user’s intentions. *See also id.*, 8:29-35 (“Natural Language Processing (NLP) involves the lexical, syntactic (grammatical), and domain semantic analysis of the user input using both statistical observations of the various surface forms and a deeper interpretation of the relationships and dependencies among words, phrases, and concepts.”), 9:39-44 (“Dialogue Manager 120 employs a series of models that help it interpret the user input and decide on the next system action.”).

122. In addition, Aretoulaki discloses selecting “*the purchase opportunity . . . based further on the interpretation of the natural language utterance*.” For example, as I discussed earlier, Aretoulaki’s processing modules output “frames” meeting the claimed “*interpretation of the natural language utterance*” and which “make[] up the so-called product (or service) description” that the Knowledge Manager receives and uses “for the dynamic database look-up.” *Id.*, 8:40-46. Based on this “product (or service) description,” the Knowledge Manager “dynamically accesses the application database 125 . . . and attempts to retrieve the best matches to the user’s requirements as specified up to that point,” after which the system delivers the results to the user as the purchase opportunity. *Id.*, 15:44-51. Thus, Aretoulaki discloses that the purchase opportunity for a particular product or service

“is selected based further on the interpretation of the natural language utterance,”

because it selects the products or services for the purchase opportunity by searching a database using the frames corresponding to the interpretation of the natural language utterance. The combination Aretoulaki and Ramer therefore renders limitation [5.b] obvious.

7. Claim 9: “The method of claim 1, the method further comprising: 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.”

123. Aretoulaki in combination with Ramer renders claim 9 obvious. First, Aretoulaki discloses *“using, by the one or more processors, an environmental model to determine environmental information.”* For example, Aretoulaki’s system maintains “a series of models that help it interpret the user input,” including a “discourse history” that “is a record of everything that has taken place up till then in the dialogue, system messages and user input, their semantic representation in terms of actions and domain parameters, and their ordering, in time.” EX1004, 9:39-10:10; *see also id.*, 15:23-25 (“discourse history” includes “system and user dialogue acts along with parameter instantiations”). Aretoulaki’s Dialogue Manager uses the discourse history to determine information of past user activity and actions, such as “user input,” “user dialogue acts,” etc., for which to interpret received utterances. *See id.*, 10:5-10, 15:23-25. Based on the ’758 patent’s explanation that

“environmental information” may correspond to “a user activity, or a user action,” a POSITA thus would have recognized that Aretoulaki’s discourse history is an “*environmental model*” that the system uses to determine “*environmental information*,” e.g., past user activities and actions as represented by user dialogue acts. See EX1001, claim 10.

124. In addition, Aretoulaki discloses that “*the context for the natural language utterance is determined based further on the environmental information.*” As discussed above, Aretoulaki discloses that the system (e.g., “*one or more processors*”) determines a context for the natural language utterance by generating “a number of different interpretations . . . representing the various alternative meanings” corresponding to different potential contexts/domains, e.g., if “the user employs a keyword that is relevant to more than a single domain or application,” and selecting an interpretation having “the highest weight.” EX1004, 15:16-29. Aretoulaki further discloses that the selection process includes weighing the different interpretations based on information of “system and user dialogue acts along with parameter instantiations” kept “in the discourse history.” *Id.*, 15:23-29. Thus, Aretoulaki discloses determining “*the context for the natural language utterance . . . based . . . on the environmental information*,” because it discloses selecting the interpretation associated with a particular context/domain based on the environmental information (e.g., user activity as indicated by “user dialogue acts”).

125. In addition, the combined system of Aretoulaki and Ramer also discloses the “*using . . . an environmental model to determine environmental information, wherein the context for the natural language utterance is determined based further on the environmental information.*” For example, Ramer discloses maintaining “mobile subscriber characteristics” that include information such as a user’s “home address,” “work address,” or other data “related to phone usage and location.” EX1005, 2:56-63, 8:57-58, 17:64-18:16. A POSITA would have understood that the mobile subscriber characteristics are an “*environmental model*” that the system uses to determine “*environmental information*,” e.g., user location information. *See also* EX1001, claim 10 (“environmental information comprises a user location”). Ramer further discloses using the user location information to generate potential interpretations of the user input. EX1005, 28:6-8 (“queries may be disambiguated or results filtered, sorted, or presented based on area code (or other location information)”). For example, when the user makes a query that includes “Royals,” the system may access the “mobile subscriber database” to determine “the user’s residence of Kansas City, Mo.,” and, based on this information of the user’s location, “predict (i.e., disambiguate) that the user is more likely seeking information pertaining to the Kansas City Royals baseball team than information about the family of Swedish Royals.” *Id.*, 24:16-27.

126. In the combined system, a POSITA would have found it obvious to modify Aretoulaki to incorporate Ramer's teaching of using user location information to interpret and disambiguate user queries. For example, as discussed above, Aretoulaki discloses determining a context for the natural language utterance by generating "a number of different interpretations . . . representing the various alternative meanings in the case of ambiguities," and selecting an interpretation having "the highest weight." EX1004, 15:16-29. Ramer further teaches using user location information to disambiguate otherwise ambiguous queries, and select certain interpretations over others. *See* EX1005, 24:16-27. A POSITA would have been motivated to modify Aretoulaki to incorporate Ramer's consideration of user location information when weighing potential interpretations of an utterance to determine the context (e.g., domain or application) associated with the utterance. This would improve Aretoulaki's system by providing additional ways to weigh potential interpretations of the user's request and clarify potential ambiguities, "speed[ing] up the interaction and the search process, [and] avoiding tedious questioning of the user." EX1004, 14:15-18. A POSITA would have found this modification obvious as a combination of prior art elements (considering discourse history when interpreting a user request, and considering user location when interpreting a user request) according to known methods to yield predictable results (considering both discourse history and user location when interpreting user

requests), and would have found the modification obvious to try as a way to create a more robust and natural interface by being able to take advantage of a greater variety of user and contextual information when interpreting received utterances. The combination Aretoulaki and Ramer therefore renders claim 9 obvious.

8. Claim 10: “The method of claim 1, wherein the environmental information comprises a user location, a user activity, or a user action.”

127. Aretoulaki in combination with Ramer renders claim 10 obvious. As discussed above, Aretoulaki discloses “*environmental information*” corresponding to previous user inputs and dialogue acts, which correspond to past “*user activity*” or “*user action*.” *See id.*, 10:5-10, 15:23-25. In addition, in the combination with Ramer discussed above, Ramer further discloses “*environmental information*” corresponding to the user’s “home address,” “work address,” or other data “related to phone usage and location.” EX1005, 2:56-63, 8:57-58, 17:64-18:16. Thus, the combination of Aretoulaki and Ramer discloses that the environmental information comprises “*a user location*” corresponding to the user’s home or work address, or a location of the user’s phone. The combination of Aretoulaki and Ramer therefore render claim 10 obvious.

9. Claim 21: “The method of claim 1, wherein the context comprises a categorization of the recognized words or phrases.”

128. Aretoulaki in combination with Ramer renders claim 21 obvious. Aretoulaki discloses that the system determines the context of the natural language

utterance by using “domain-specific grammars” and a “mapping lexicon” to identify “individual words” of the utterance “point[ing] to domain specifications,” and translating “application-specific words into the corresponding relevant concepts in the domain ontology.” EX1004, 4:48, 11:41-45; *see also id.*, 15:19-22 (describing that ambiguity may arise when the recognized words or phrases include keywords that are “relevant to more than a single domain or application”). Aretoulaki thus discloses that the determined context comprises “*categorization of the recognized words or phrases*” because it categorizes the words or phrases as being associated with a specific application or domain using “domain-specific grammars” and mapping them to a specific domain ontology. The combination of Aretoulaki and Ramer therefore renders claim 21 obvious.

10. Claim 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.”

129. Aretoulaki in combination with Ramer renders claim 22 obvious. Aretoulaki discloses that the context of the natural language corresponds to the domain or application associated with the utterance, which may include domains or applications for travel booking and pricing (*See* EX1004, 15:20-22), purchasing electronics such as printers or scanners (*see id.*, 20:29-34), etc. Thus, Aretoulaki discloses at least that the context comprises one of “*e-commerce*” (e.g., travel

booking and pricing, electronics purchasing, etc.) or “*transportation*” (e.g., travel booking and pricing). The combination of Aretoulaki and Ramer therefore renders claim 22 obvious.

11. Claim 23

130. Claim 23 is rendered obvious for similar reasons as claim 1. The claimed “*system*” is met by Aretoulaki’s “Knowledge Communications Platform,” which includes various processing modules such as the “Communications Mediator,” “NLP Manager,” “Knowledge Manager,” and “Dialogue Manager.” EX1004, 7:50-8:8, FIG. 1. As discussed above, these components perform “*processing [of] natural language utterances that include requests and selecting and presenting purchase opportunities based thereon,*” and a POSITA would have understood that they include “*one or more physical processors programmed with computer instructions, which when executed cause the one or more physical processors to*” perform certain steps. *See id.* The remaining elements of claim 23 are substantively identical to those of claim 1.

12. Claim 24

131. Claim 24 is substantively identical to claims 2 and is obvious for the same reasons.

13. Claim 25

132. Claim 25 is substantively identical to claim 3 and is obvious for the same reasons.

14. Claim 26

133. Claim 26 is substantively identical to claim 4 and is obvious for the same reasons.

15. Claim 27

134. Claim 27 is substantively identical to claim 5 and is obvious for the same reasons.

16. Claim 31

135. Claim 31 is substantively identical to claim 9 and is obvious for the same reasons.

17. Claim 32

136. Claim 32 is substantively identical to claim 10 and is obvious for the same reasons.

18. Claim 43

137. Claim 43 is substantively identical to claim 21 and is obvious for the same reasons.

19. Claim 44

138. Claim 44 is substantively identical to claim 22 and is obvious for the same reasons.

B. Ground 2: Aretoulaki in combination with Ramer and Jong render obvious claims 8 and 30

1. Claim 8: “The method of claim 1, wherein recognizing the words or phrases from the natural language utterance comprises: 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.”

139. Aretoulaki in combination with Ramer and Jong renders claim 8 obvious. As discussed above for limitations [1.a] and [1.b], Aretoulaki and Ramer disclose performing speech recognition to recognize the words or phrases of the natural language utterance. For example, as discussed above, Aretoulaki’s “Communications Mediator 105” “handles input and output of different modalities,” including “speech,” and performs speech recognition to convert “the user input into text form.” EX1004, 8:9-16. Ramer also discloses a “voice recognition facility” that converts received human voice content into recognized words “through the use of a digital database storing a vocabulary of words or syllables, coupled with a means of comparing this stored data with the digital voice signals received by the device.” EX1005, 21:25-36.

140. To the extent that Aretoulaki and Ramer leave open the specific techniques used to recognize speech, and do not explicitly disclose that the speech recognition engine uses “*phonetic dictation to recognize a stream of phonemes in the natural language utterance, where the recognized words or phrases are based on the recognized stream of phonemes,*” the combination with Jong renders it

obvious. Jong discloses a “speech recognition device” that performs speech recognition by translating received “speech input signals into textual representations . . . representing the phonemes detected in the speech input signals.” EX1008, Abstract, 9:34-40. The speech recognition device detects the phonemes using a “word model device” that “includes a listing of phonemes and corresponding textual code representations of these phonemes.” *Id.*, 9:41-46. A POSITA would have understood that the recognized phonemes may be “*a stream of phonemes*,” because Jong discloses transmitting text data derived from the phonemes as “a continuous stream.” *Id.*, Abstract. A POSITA would also have understood that identifying the phonemes in the speech signal using a listing mapping phonemes to “corresponding textual code representations” corresponds to using “*phonetic dictation to recognize [the] stream of phonemes in the natural language utterance*,” because phonetic dictation involves representing the speech sounds (phonetics) of a received speech signal. *See id.*, 9:41-46.

141. Jong further discloses that the “*recognized words or phrases [of the natural language utterance] are based on the recognized stream of phonemes*.” For example, Jong discloses that after identifying the stream of phonemes of the received natural language utterance, the speech recognition device uses a “subword model database 304 contain[ing] word syllables that are correlated with the phonemes of the word model device 303” to extract “corresponding syllables” from

the phonemes, and checks the determined syllables “against the lexicon data stored in [a] lexicon database” to group the syllables into “recognizable words.” EX1008, 5:50-67. Thus, Jong discloses that the “recognizable words” of the speech signal (“*recognized words or phrases*”) are based on the recognized stream of phonemes, being grouped from syllables identified from the recognized phonemes. *See also id.*, 10:57-62 (claim 8, “determining phonemes included in the first speech input signals and converting the phonemes into . . . textual words”).

142. Aretoulaki, Ramer, and Jong are analogous, all describing systems that perform speech recognition. *See, e.g.*, EX1004, 8:12-18 (describing translating received speech queries into an “input text string”); EX1005, 21:25-36 (describing “voice recognition” to convert “human voice” data into recognized text using “a vocabulary of words or syllables”); EX1008, 5:50-67 (converting speech signals to phonemes to syllables to words using subword models and a stored lexicon). While Aretoulaki-Ramer leaves open the specific technique it uses to recognize speech, Jong provides a specific speech recognition technique that a POSITA would have recognized as straightforward to implement in Aretoulaki-Ramer’s system. A POSITA would thus have been motivated to modify the speech recognition engine of Aretoulaki-Ramer to perform phonetic dictation to recognize a stream of phonemes in the natural language utterance as taught by Jong, which are then

translated into syllables and into words, as a specific way for the system to perform speech recognition to recognize words and phrases from a user's voice input.

143. The modification is a combination of prior art elements disclosed in Aretoulaki, Ramer, and Jong according to known methods to yield predictable results. A POSITA would have had a reasonable expectation of success because the combination is a simple substitution of a specific known technique for speech recognition (e.g., Jong's phonetic dictation that recognizes phonemes in a natural language utterance, and recognizes words based on the phonemes) for more the general disclosure of speech recognition found in Aretoulaki-Ramer to yield predictable results, e.g., Aretoulaki-Ramer performing speech recognition using Jong's specific technique. A POSITA would have also found the modification obvious to try as there were a finite number of well-known ways to perform speech recognition, and would have had a reasonable expectation of success that the speech recognition performed in Aretoulaki-Ramer could be done by using phonetic dictation to recognize phonemes as taught in Jong. The combination of Aretoulaki, Ramer, and Jong therefore renders claim 8 obvious.

2. Claim 30

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

C. Ground 3: Aretoulaki in combination with Ramer and Kennewick render obvious claims 6-7, 11-20, 28-29, and 33-42

1. General Motivations to Combine Aretoulaki, Ramer, and Kennewick

145. A POSITA would have found it obvious to combine Aretoulaki and Ramer with Kennewick. These references are all in the same field of interactive speech technology, all describing systems and methods for enabling users to conduct voice-based e-commerce to search for and purchase goods and services through an interactive speech interface. For example, as discussed above, Aretoulaki discloses a system and method for “intuitive natural language (text or speech) dialogue-based electronic commercial transactions and information exchanges,” which processes natural language queries to select purchasable products or services for presentation to the user. EX1004, Abstract, 8:27-48, 22:10-23. Ramer discloses a “wireless search platform [] architecture” that “facilitates the processing of user-initiated [voice] queries” for purchasable products or services. EX1005, 4:7-10, 15:47-49, 47:34-39, 47:63-48:12. Kennewick similarly discloses a “natural language speech system” that “receives spoken natural language queries, commands and/or other utterances from a user . . . processes the input . . . and responds to the user with a natural language speech response.” EX1006, [0012], [0014]. Kennewick’s “speech interface device” also provides “offers and promotions for goods and services”

responsive to user queries and can facilitate “remote ordering” for goods and services. *Id.*, [0066]-[0067].

146. Aretoulaki, Ramer, and Kennewick also share similar architectures for an interactive speech system. Each discloses an input device that captures a user’s voice input. *See, e.g.*, EX1004, Abstract (user can “pose questions over the Internet” through “a PC microphone, a fixed or mobile phone,” etc.); EX1005, 6:21-29 (“The voice entry 122 function . . . may be used through the speaker-receiver device of the mobile communication facility 102”); EX1006, [0121] (“speech unit 128 includes one or more microphone . . . to receive the utterances of the user”), claim 1 (“a speech unit that receives the user generated natural language speech utterance”). Each discloses selecting and presenting a product or service to the user’s request based on the user’s voice input. *See, e.g.*, EX1004, 8:27-48 (processing “user queries” to determine a “product (or service) description which will be used for the dynamic database look-up”), 15:46-51 (presenting a “result set” of products or services “to the user . . . as either A[S]CII text, speech and/or images”); EX1005, 47:63-48:3 (“presenting a link in a user interface... to an item of goods or services for sale”); EX1006, [0028] (“parsing and interpreting the question or query”), [0066]-[0067] (“provid[ing] interactive offers and promotions for goods and services” based on user queries). Each discloses enabling the user to purchase presented goods or services through the interactive speech system. *See, e.g.*, EX1004, 22:20-26 (“user

can select” a product or service “proposed by the system and pose additional questions . . . and/or ask to make a purchase (commit)”); EX1005, 48:4-15 (presented link includes “various e-commerce features, such as those enabling single click purchasing, . . . instant purchasing (e.g., ‘buy it now’)”); EX1006, [0018] (describing “fast food ordering,” “travel services,” etc.), [0067] (“The system facilitates remote ordering” and “payment”).

147. Additionally, Aretoulaki and Kennewick also both teach processing and responding to user utterances relating to different domains or applications using domain-specific information, by determining a context or domain associated with the utterance. Aretoulaki discloses interpreting user inputs relating to different domains and applications by accessing domain-specific knowledge and concepts. *See, e.g.*, EX1004, 4:37-42 (“active look-up of a knowledge base with expert domain knowledge in the course of the dialogue in order to identify and correct early mismatches between the user's beliefs and the stored system data about domain entities”), 8:50-52 (“Knowledge Manager 115 maintains and manipulates information on both the world in general and the specific domain and application under consideration.”), 10:15-18 (“the close integration of knowledge about the application domain and the world in the interpretation process,” as coordinated by the Dialogue Manager). For example, Aretoulaki performs Natural Language Processing (NLP) on received user utterances to generate a “domain semantic

representation 225” of utterance indicating “the desired product, service or information,” by identifying “individual words” “point[ing] to domain specifications,” and using “a mapping lexicon [to] translate[] application-specific words into the corresponding relevant concepts in the domain ontology.” *Id.*, 11:31-45. The Dialogue Manager may perform further “discourse analysis” on the “initial interpretation” using domain-specific data accessed through the Knowledge Manager to generate a “domain and dialogue act semantic representation” of the utterance establishing “the salient domain concepts and their values appearing in the user input.” *Id.*, 8:50-52, 12:15-19, 15:44-45. The Knowledge Manager, which is responsible for “maintain[ing] and manipul[at]ing information” of “the specific domain and application under consideration,” searches “the application database” based on the generated interpretation to select the “best matches to the user’s requirements.” *Id.*, 8:50-52, 12:15-19, 15:46-51.

148. Kennewick similarly also teaches processing user “queries or commands in multiple domains.” EX1006, [0011]. To do so, Kennewick discloses determining a “most[] likely context or domain for a user’s question or command,” e.g., by analyzing the recognized words or phrase of the utterance and performing “keyword matching” to determine the context. *Id.*, [0017], [0160], [0161]. Based on the determined context, the parser “determine[s] the domain and thereby, the domain agent 156, if any, to be invoked,” where the domain agent manages the

domain specific behavior and information for the specific domain. *Id.*, [0032], [0160]; *see also id.*, [0126] (The domain agent “include[s] everything needed to extend or modify the functionality of the system in a current or new domain”). The invoked domain agent “make[s] one or more queries and may rapidly return a formatted response” based on the user request. *Id.*, [0032], [0167].

149. A POSITA in possession of Aretoulaki, Ramer, and Kennewick would have been motivated to modify Aretoulaki’s system to incorporate Kennewick’s teachings of using domain agents for managing domain-specific behaviors and information for different domains. For example, Aretoulaki’s system includes a Knowledge Manager responsible for “maintain[ing] and manipulat[ing] information on . . . specific domain[s] and application[s] under consideration” and accessing “the most current data from resource agent 125 (i.e., application databases, a product catalogue, ontology, etc.).” EX1004, 8:50-52, 9:13-16. A POSITA would have found it obvious to modify Aretoulaki’s Knowledge Manager to maintain and manipulate information of different domains and applications through an agent architecture containing domain agents as taught Kennewick, to create a more modular system that can be more easily extended into new domains and applications. *See* EX1006, [0126] (The domain agent “include[s] everything needed to extend or modify the functionality of the system in a current or new domain”). Indeed, Aretoulaki already discloses that the Knowledge Manager accesses data associated

with different applications and domains (e.g., application databases) through a “resource agent,” and it would have been obvious to organize the knowledge manager and/or the resource agent into different application- or domain-specific agents as taught by Kennewick, where each domain agent manages the resources for a respective domain. *See* EX1004, 9:12-17.

150. The modification is a combination of prior art elements disclosed in Aretoulaki and Kennewick according to known methods to yield predictable results. 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 different domains through respective domain agents) for another (managing domain-specific behavior and data for different domains through a Knowledge Manager) to obtain predictable results (managing domain-specific behavior and data for different domains through a Knowledge Manager that uses respective domain agents). 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 as a way to create a more robust and natural interface. I provide additional motivations to combine the references below.

2. Claim 6: “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: (i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.”

151. Aretoulaki in combination with Ramer and Kennewick renders claim 6 obvious. As discussed above, Aretoulaki discloses a “*conversational language processor*” that includes an “NLP Manager,” “Dialogue Manager,” and “Knowledge Manager” which interpret the natural language utterance using domain-specific information to “*interpret[]*” and “*determin[e] a response to the requests related to [a] domain.*” See EX1004, 8:4-7. For example, Aretoulaki’s “*conversational language processor*” uses “domain-specific grammars and lexica” to determine the domain or application associated with the utterance. *Id.*, 4:48-51. Based on the determined domain/application, the Knowledge Manager, which “maintains and manipulates information on . . . the specific domain and application under consideration,” accesses domain-specific information provided by a “resource agent,” including “application databases, a product catalogue, ontology, etc.” *Id.*, 8:50-52, 9:13-16. The Knowledge Manager communicates this information to the Dialogue Manager to assist in interpreting the user utterance and deciding “on the next system action.” *Id.*, 9:16-20; see also *id.*, 13:37-45 (The NLP Manager 110 and/or Knowledge Manager 115 isolate or infer “application-specific parameters” and “domain concepts” from the user input). Thus, Aretoulaki discloses that the

Knowledge Manager uses domain-specific information for a particular domain to “*assist in . . . interpreting requests related to its domain.*”

152. Aretoulaki further discloses utilizing domain-specific information for a domain for “*determining a response to the requests related to its domain.*” For example, Aretoulaki discloses that after the system establishes “the salient domain concepts and their values appearing in the user input,” the Knowledge Manager “dynamically accesses the application database 125 . . . and attempts to retrieve the best matches to the user’s requirements.” *Id.*, 15:44-49. The system uses the retrieved “Result Set” to create a response for presentation to the user. *Id.*, 15:50-51. Thus, Aretoulaki discloses the Knowledge Manager uses domain-specific information for a particular domain for “*determining a response to the requests related to the domain.*”

153. To the extent that Aretoulaki does not explicitly disclose the conversational language processor, including the Knowledge Manager, comprising “*one or more domain agents,*” the combination with Kennewick renders it obvious. As I discuss in greater detail below, Kennewick discloses a “*conversational language processor*” configured to interpret a natural language utterance that includes a “parser 118” and “agents 106.” *See, e.g.,* EX1006, [0123] (parser interprets the utterance “using data supplied by the agents”), [0163] (describing that the “parser and domain agent use” domain-specific information “to determine the

most plausible interpretation” of an utterance). Kennewick further discloses that the “agents 106” include “*one or more domain agents*,” where each domain agent includes “the behavior and data required for a specific area of functionality” for assisting the parser in interpreting requests relating to its domain. *Id.*, [0018], [0160]. For example, FIG. 6, reproduced below, shows that the agents 106 include multiple domain agents 156.

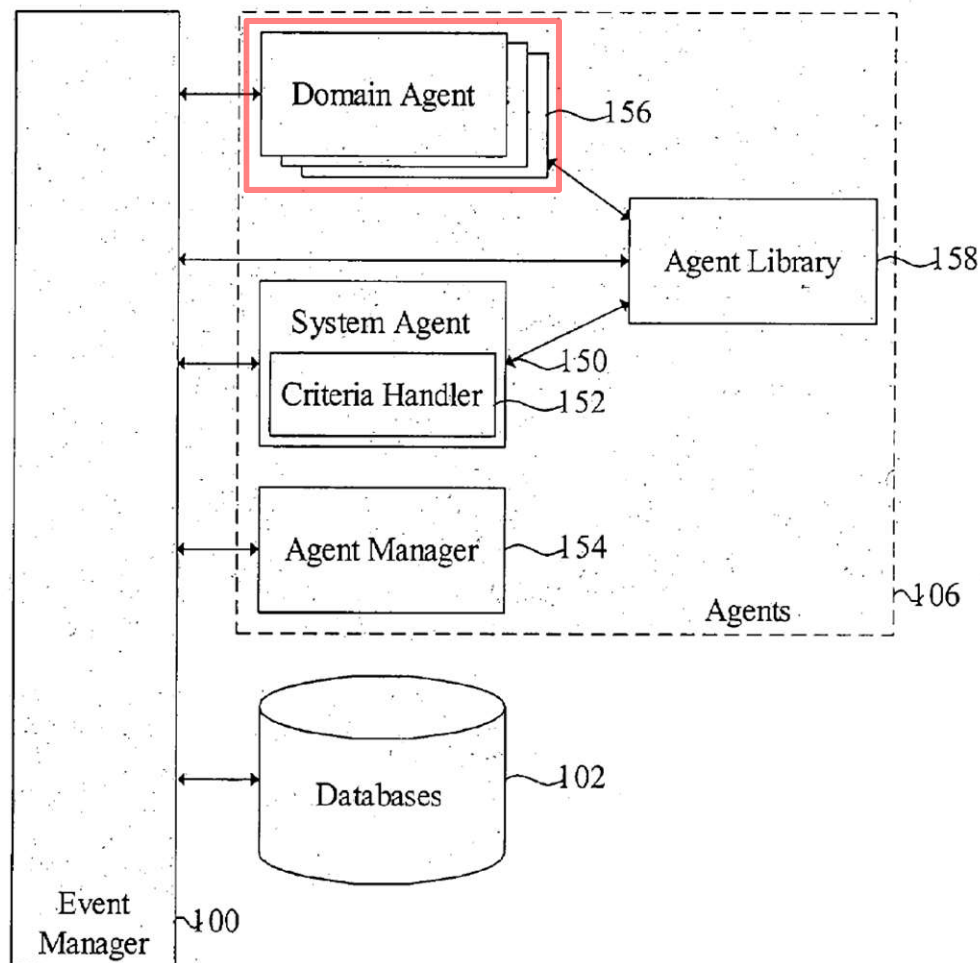


Figure 6. Agent Architecture

154. Kennewick discloses that each domain agent includes “the behavior and data required for a specific area of functionality” for assisting the parser in interpreting requests relating to its domain. EX1006, [0018]. For example, the “parser and domain agent [may] use” domain-specific information “to determine the most plausible interpretation” of an utterance and resolve potential ambiguities. *Id.*, [0163]; *see also id.*, [0029] (interpreting an utterance “may be performed using . . . domain specific information”). Thus, Kennewick discloses that the domain agent is configured to assist in “*interpreting requests related to its domain*” because it provides data to the parser for interpreting user requests related to its domain. *See also id.*, [0163] (describing a travel services domain agent using “flight information in the database and network information” to assist the parser in interpreting an utterance relating to a flights domain). In addition, each domain agent further assists in “*determining a response to the request relating to its domain*” by “create[ing] one or more queries to one or more local or external information sources” related to its domain, and “create[ing] a satisfactory response to [the] question, or to [the] command.” *Id.*, [0167], [0178].

155. A POSITA would have found it obvious to modify Aretoulaki based on the teaching of Kennewick, where in the combined system, Aretoulaki’s Knowledge Manager, to more efficiently “maintain[] and manipulate[] information on . . . specific domain[s] and application[s] under consideration,” is modified to utilize an

agent architecture containing domain agents configured to maintain and manipulate information for respective domains and applications. *See* EX1004, 8:50-52. For example, a POSITA would have found it obvious to organize the functionality of the Knowledge Manager of managing information for different domains into different agents that each manage information for a respective domain. *See* EX1006, [0017] (“domain specific behavior and information are organized into agents”). A POSITA would have been motivated to do so to create a more modular system that can be more easily extended into new domains and to incorporate new applications, e.g., by adding a new agent to the Knowledge Manager. *See* EX1006, [0126] (“an agent may include everything needed to extend or modify the functionality of the system in a current or new domain”). Indeed, Aretoulaki already discloses that the Knowledge Manager accesses data associated with 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 application- or domain-specific agents as taught by Kennewick. EX1004, 9:12-17.

156. In my opinion, a POSITA would therefore have been motivated to combine Aretoulaki, Ramer and Kennewick. As discussed above, the references are analogous, all relating to natural language interfaces through which users can conduct e-commerce by purchasing products or services using natural language utterances. *See, e.g.,* EX1004, Abstract, 8:27-48, 11:31-36, 22:10-23; EX1005,

4:7-10, 15:47-49, 47:34-39, 47:63-48:13; EX1006, Abstract, [0018], [0060]-[0068].

The combination is a simple substitution of one known technique (managing domain-specific behavior and data for different domains through a Knowledge Manager) for another (managing domain-specific behavior and data for different domains through respective domain agents) to yield predictable results, and a POSITA would have been motivated to make this combination as it provides a straightforward way to make Aretoulaki's system extendible to incorporate data and functionality for additional applications and domains. The combination of Aretoulaki, Ramer, and Kennewick thus renders claim 6 obvious.

3. Claim 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.”

157. Aretoulaki in combination with Ramer and Kennewick renders claim 7 obvious. For example, Aretoulaki's system is configured to “provid[e] a natural language dialogue-based electronic commercial transaction” and supports different domains including “*electronic commerce domain[s]*” through which the user can purchase goods or services such as travel packages, electronics, etc. EX1004, 7:50-52, 15:20-22, 20:29-34, 22:10-18 (describing linking user to “E-commerce sites”).

158. In addition, as discussed above, Kennewick discloses “*one or more domain agents*” which each “include the behavior and data required for a specific

area of functionality” or “specific domain[.]” EX1006, [0018], [0127]. These include various “*electronic commerce domain*[s]” for “travel services, entertainment scheduling,” and “remote ordering and payment for goods and services” (e.g., “fast food ordering”). *Id.*, [0018], [0067]. Thus, Kennewick discloses that the domains in which the one or more domain agents are configured to assist comprises an electronic commerce domain.

159. In the combined system discussed above in which Aretoulaki’s Knowledge Manager is modified to manage interactions with different applications and domains through domain agents as taught by Kennewick, the domains may include electronic commerce domains as taught by Aretoulaki and Kennewick. For example, a POSITA would have found it obvious for Aretoulaki’s Knowledge Manager, in the combined system, to utilize a travel services domain agent as taught by Kennewick to support an application for booking travel arrangements, a different domain agent for supporting an application for purchasing printers and other electronics, etc. *See* EX1004, 15:53-54 (describing applications for purchasing “printers” and for booking “travel offers”), 16:31-34; EX1006, [0018] (“Examples of query and response domains include . . . travel services”). The combination of Aretoulaki, Ramer, and Kennewick thus renders claim 7 obvious.

4. Claim 11

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

160. Aretoulaki in combination with Ramer renders obvious the method of claim 1.

[11.a] “invoking, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;”

161. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. As discussed above for limitation [1.d], Aretoulaki discloses that the system’s processing modules (e.g., “***one or more processors***”) “***select the purchase opportunity based on the determined context.***” For example, the Knowledge Manager receives domain semantic representation of “the user’s requirements . . . about the desired product, service or information” generated based on the determined context, and uses it to “access[] [an] application database,” e.g., via a “resource agent,” to retrieve “the best matches to the user’s requirements” and select the purchase opportunity to be delivered to the user. EX1004, 9:12-17, 11:31-40, 15:44-51. A POSITA would have understood that the Knowledge Manager, resource agent, and/or application database are examples of “***system resources***” invoked “***to select the purchase opportunity based on the determined context,***” being resources that are available to the system and used to select the purchase opportunity based on the determined context, e.g., based on the “domain semantic representation.” Aretoulaki therefore renders this limitation obvious.

162. In addition, the combination with Kennewick also renders the limitation obvious. As discussed above, Kennewick discloses “invoking the proper resources, including agents” as part of processing each user query or command. EX1006, [0017], [0028]. A POSITA would have understood that the invoked resources including agents are “**system resources**,” being resources that are available for use by the system. *See also id.*, FIG. 5 (illustrating the speech recognition engine, parser, and agents are resources that can be invoked by the speech processing unit 98). Thus, Kennewick discloses “**invoking . . . one or more system resources**” to process and respond to the natural language utterance.

163. In the combined system, a POSITA would have found it obvious to invoke system resources including the appropriate domain agent to “***select the purchase opportunity based on the determined context.***” *See* EX1006, [0162]. For example, a POSITA would have found it obvious that in the combined system where Aretoulaki’s Knowledge Manager is modified to utilize an agent architecture including domain agents as taught by Kennewick, the Knowledge Manager would invoke an appropriate agent based on the determined context, which would select the purchase opportunity by accessing the appropriate application database as taught by Aretoulaki, resulting in the purchase opportunity being selected based on the determined context. The combination of Aretoulaki, Ramer, and Kennewick therefore renders this limitation obvious.

[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.”

164. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. As discussed above, a POSITA would have found it obvious for the combined system to invoke “***system resources [that] include at least an agent,***” e.g., a domain agent corresponding to the determined context as taught in Kennewick, to select the purchase opportunity based on the determined context. Each domain agent is “***an autonomous executable***” comprising a “convenient and ***re-distributable package[] . . . of functionality***” for “a specific domain” or “application area,” that “***receives, processes and responds to***” received user questions, queries, and commands pertaining to its domain. EX1006, claim 1, [0017] (agents are “packages of executable code, scripts, links to information, data, and other data forms”), [0126]. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [11.b] obvious.

5. Claim 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.”

165. Aretoulaki in combination with Ramer and Kennewick renders claim 12 obvious. First, Aretoulaki discloses returning “***results of user questions, queries, and/or commands as responses to the user.***” For example, as discussed above,

Aretoulaki discloses receiving “user queries” via the Communications Mediator, and interpreting the queries to generate “a representation of the user's requirements as a domain semantic representation 225 (or some of these requirements) about the desired product, service or information.” EX1004, 8:1-16, 11:31-40. The Knowledge Manager acquires the “domain semantic representation 225 . . . to be matched against the descriptions of existing products, services and information in the application databases,” and retrieves “the best matches to the user’s requirements as specified up to that point.” *Id.*, 11:31-40, 15:44-51. The Knowledge Manager returns the retrieved “result set” as a response to the user via the Dialog Manager and Communications Mediator “as either A[S]CII text, speech and/or images.” *Id.*, 15:44-51. Thus, the Knowledge Manager “*returns results of user questions, queries, and/or commands as responses to the user.*”

166. The Knowledge Manager matches the domain semantic representation against information in the application database by querying the application database to produce the “result set.” *See* EX1004, 15:44-51, 16:5-10 (“databases search”), 24:12-13 (“database query”). Aretoulaki thus discloses creating the response using “*the results of information queries*,” e.g., the “result set.” In addition, Aretoulaki discloses creating the response based on “*user preferences or other data in the user-specific profile.*” For example, the user model contains data of user preferences such as “application requirements, such as favoring Mediterranean destinations, and

presentation preferences, e.g. images only,” which Aretoulaki’s system uses to configure the content of the response (e.g., Mediterranean destinations) and/or the manner in which the response is conveyed (e.g., images only). *Id.*, 9:47-52. Thus, Aretoulaki discloses returning the “*results of user questions, queries, and/or commands as responses to the user*,” which are created using the “*results of information queries*” (e.g., the “result set”), and “*user preferences or other data in the user-specific profile*.”

167. To the extent that Aretoulaki does not explicitly disclose “*the agent*” returning the “*results of user questions, queries, and/or commands as responses to the user*,” and creating the response using a “*using the results of information queries, system personality, and user preferences or other data in the user-specific profile*,” the combination with Kennewick renders it obvious. For example, Kennewick discloses that the “*agent*” “*returns results of user questions*” as “*responses to users*,” where the response is 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]. For example, the invoked agent is configured to “receive, process and respond to user questions, queries and commands” by “creat[ing] one or more queries to one or more local or external information sources” and “evaluat[ing] the results of the one or more queries” with which to construct “a satisfactory response” to the user. *Id.*, [0017],

[0167], [0170]-[0171], [0175]. In addition, the agent, when creating the response, “select[s] the presentation[/system] personality 108 to be used,” and “appl[ies] a number of transformations to the tokens before presentation to the user” that are “derived from a number of sources including . . . the user profile 110.” *Id.*, [0178], [0180]; *see also id.*, [0156] (“system personality 108”). A POSITA would have understood that in the combined system in which Aretoulaki’s Knowledge Manager invokes an appropriate agent based on the selected domain or application, the invoked agent would create the responses in the same way as taught by Kennewick, resulting in a response that is “*created using the results of information queries, system personality, and user preferences or other data in the user-specific profile.*” The combination of Aretoulaki, Ramer, and Kennewick therefore renders claim 12 obvious.

6. Claim 13: “The method of claim 11, further comprising updating the agent from a remote location over a network to provide new functionality.”

168. Aretoulaki in combination with Ramer and Kennewick renders claim 13 obvious. For example, as discussed above, Kennewick discloses “*updating the agent from a remote location over a network,*” e.g., the Internet, using an “update manager” that “manages the automatic and manual loading and updating of agents and their associated data from the Internet.” EX1006, [0120]; *see also id.*, [0017], [0126]. Kennewick further discloses updating the agent to “*provide new*

functionality” because Kennewick discloses updating an agent to “add[] new behaviors,” which a POSITA would have understood may correspond to “*new functionality*.” *Id.*, [0017], [0019], [0126]; *see also id.*, [0129] (“[d]evelopers of agents 106 can add new functionality to the agent library 158 as required”).

169. In the combined system in which Aretoulaki’s Knowledge Manager is modified to utilize Kennewick’s agent architecture having redistributable agents for supporting different areas of functionality and accessing domain-specific data, a POSITA would have found it obvious to update the agents from a remote location to provide new functionality in the same way as taught by Kennewick. A POSITA would have been motivated to make this modification to improve Aretoulaki’s system by making it easier to extend the system with new domains and applications through the use of agents, and to update the agents with new functionality when available. The combination of Aretoulaki, Ramer, and Kennewick therefore renders claim 13 obvious.

7. Claim 14: “The method of claim 11, wherein the agent uses system resources and the services of other agents.”

170. Aretoulaki in combination with Ramer and Kennewick renders claim 14 obvious. In the combined system in which Aretoulaki’s Knowledge Manager is modified to utilize Kennewick’s agent architecture, Kennewick discloses that the agent uses “*system resources and the services of other*, typically more specialized, *agents*.” EX1006, [0017].

8. Claim 15: “The method of claim 11, further comprising providing a framework allowing the user to create or modify the agent.”

171. Aretoulaki in combination with Ramer and Kennewick renders claim 15 obvious. In the combined system in which Aretoulaki’s Knowledge Manager is modified to utilize Kennewick’s agent architecture, Kennewick discloses providing “*a framework*” such as a speech interface or graphical user interface through which the user can “create and manage modifications to [the] agent[.]” EX1006, [0134]. A POSITA would have found it obvious for the combined system to implement this framework as taught by Kennewick, to enable users to “extend the system and configure it to their own preferences.” EX1006, [0021].

9. Claim 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.”

172. Aretoulaki in combination with Ramer and Kennewick render claim 16 obvious. In the combined system in which Aretoulaki’s Knowledge Manager is modified to utilize Kennewick’s agent architecture, Kennewick discloses “*providing a license management framework*” because it discloses an “agent manager 154” that provides a “license management client” that provide “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] (available license management schemes include “out right purchase, subscription for updates, one time or limited time use”). *Id.* Thus, Kennewick discloses providing a license management

framework to enable a third party to sell an agent, and for the user to purchase the agent, on a one time or subscription basis. A POSITA would have found it obvious for the combined system to implement a license management framework in the same way as taught by Kennewick, to allow a user to “to extend the functionality of the invention to a new domain” by licensing new domain agents. *Id.*, [0017].

10. Claim 17

[17.pre] “The method of claim 11, wherein:”

173. Aretoulaki in combination with Ramer and Kennewick renders the method of claim 11 obvious.

[17.a] “the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories the user can interact with, and receive the same services and applications from, any one of the devices on the network;”

174. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. First, Aretoulaki discloses “*a common base of agents, data, information, user-specific profiles, and histories.*” For example, Aretoulaki discloses a system (“Knowledge Communications Platform”) that includes a number of processing modules and components, including a “Knowledge Manager,” “Dialogue Manager,” and “resource agent” (“*agents*”), which utilize a common “knowledge base” containing “expert domain knowledge,” concepts, and ontologies (“*data, information*”), user models (“*user-specific profiles*”) “contain[ing] data about the general preferences and assumptions of individual users,” and discourse

histories that “record . . . everything that has taken place up till then in the dialogue” (“*histories*”). See, e.g., EX1004, 4:17-42, 7:11-45, 8:1-8, 9:6-49, 10:5-10, 14:38-45. A POSITA would have understood these components as being a “*a common base of agents, data, information, user-specific profiles, and histories.*”

175. In addition, Aretoulaki discloses that “*the method is implemented on a network of devices using [the] common base of agents, data, information, user-specific profiles, and histories,*” where the user “*the user can interact with, and receive the same services and applications from, any one of the devices on the network.*” For example, Aretoulaki discloses that a user can access the system (“Knowledge Communications Platform”) over the Internet via any one of a number of different devices, such as “a PC Web or E-mail Browser, a PC microphone, a fixed or mobile phone, or any wireless device such as a Personal Digital Assistant.” EX1004, Abstract, 3:51-4:7. A POSITA would have understood these different devices connected to a network (e.g., the Internet) are “*a network of devices.*” The user can use any of these devices to interact with the components of Aretoulaki’s system over the network to make use of “*common base of agents, data, information, user-specific profiles, and histories,*” and receive the “*same services and applications*” provided by the system, because “the user can employ any input device, as long as this input acquires some basic textual semantic representation.” *Id.*, 10:31-33. This is consistent with the disclosure of the ’758 patent, which

explains that “system 100 may receive a user input, including at least a voice-based user utterance, at an input device 105” that “may include any suitable device, or combination of devices, for receiving a voice-based input (e.g., a microphone).” EX1001, 4:7-11; *see also id.*, 5:36-53 (incorporating by reference U.S. patent application Ser. No. 11/200,164, which describes “a network of devices” as “multiple devices on which the invention is deployed, and connected to a network, . . . placed at different locations”).

176. In addition, to the extent that Aretoulaki does not explicitly disclose a network of devices “*using a common base of agents, data, information, user-specific profiles, and histories the user can interact with, and receive the same services and applications from, any one of the devices on the network,*” the combination with Kennewick renders it obvious. For example, Kennewick discloses that its system implementing its methods “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. The network of devices are “connected to a network” and “placed at different locations throughout a home, place of business, vehicle or other location,” so that the user can “*interact with, and receive the same services and applications* at any location equipped with the required device on the network.” EX1006, [0016]. Thus, Kennewick discloses that the user can interact

with, and receive the same services and applications from, “*any one of the devices on the network.*”

177. In the combination of Aretoulaki, Ramer, and Kennewick discussed above, a POSITA would have found it obvious to modify Aretoulaki, to the extent it does not explicitly disclose it, to implement its system (e.g., the “Knowledge Communications Platform”) performing the claimed method on “*a network of devices,*” where the user can “*receive the same services and applications from, any one of the devices on the network*” as taught by Kennewick. This would increase the accessibility and flexibility of the system by enabling the user to access the same system functionality using different devices at different locations, and is consistent with Aretoulaki’s disclosure of the user being able to “employ any input device” to interact with the system. EX1004, 10:31-33. In addition, in the combined system, because the user accesses the same underlying system through any of the network of devices, a POSITA would have understood that the network of devices uses a “*common base of agents, data, information, user profiles and histories.*” EX1006, [0016]. A POSITA would have found this combination obvious, as it is a simple substitution of one known element (e.g., accessing a system through a client device) for another (e.g., accessing the system through any of a network of devices) to obtain predictable results. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [17.a] obvious.

[17.b] “the location of a particular device addressed by the user is used as part of context for questions asked of the user.”

178. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. For example, in the combination discussed above, the method is implemented on a system deployed on “*a network of devices*” as taught by Kennewick. EX1006, [0016]. As disclosed in Kennewick, these devices may be “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. *Id.* A POSITA would have found it obvious for the combined system to use the location information as part of the context in the same way as taught by Kennewick, and found this modification obvious as a combination of prior art elements (considering discourse history when interpreting a user request, and considering user location when interpreting a user request) according to known methods to yield predictable results (considering discourse history and user location when interpreting user requests). A POSITA would have found the modification obvious to try as a way to create a more robust and natural interface by being able to take advantage of a greater variety of user and contextual information when interpreting received utterances. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [17.b] obvious.

11. Claim 18

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

179. Aretoulaki in combination with Ramer and Kennewick renders the preamble obvious. As discussed above, Aretoulaki in combination with Ramer renders the method of claim 1 obvious.

[18.a] “capturing, by the one or more processors, the first natural language utterance;”

180. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. The claimed term “the first natural language utterance” lacks antecedent basis. For my analysis, I assume “the first natural language utterance” in this limitation refers to the “natural language utterance” of limitation [1.a]. For example, Aretoulaki discloses that the Communications Mediator (of the “***one or more processors***”) “handles input and output of different modalities,” including speech, and translates the received speech into text, e.g., “an ASCII string.” EX1004, 8:9-18. Thus, Aretoulaki’s Communications Mediator includes a speech recognizer that “***captur[es] . . . the first natural language utterance***” for conversion into text.

181. In addition, Kennewick also discloses this limitation. For example, Kennewick discloses “a speech recognition module for ***capturing*** the user input” received as “spoken natural language queries, commands and/or other utterances from a user,” which a POSITA would have found obvious includes “***the first natural***

language utterance.” EX1006, [0013], [0014]; *see also id.*, [0028] (“For each user utterance . . . capturing the user's question or query through accurate speech recognition”). A POSITA would have understood that in the combined system discussed above, to the extent that Aretoulaki does not explicitly disclose “*capturing*” the first natural language utterance, it would be obvious that Aretoulaki’s Communications Mediator does so when it receives the user’s input, based on the teachings of Kennewick. *See* EX1004, 45:24-27. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [18.a] obvious.

[18.b] “parsing, by the one or more processors, the first natural language utterance;”

182. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. For example, Aretoulaki discloses that the system’s processing modules (e.g., the “*one or more processors*”) includes “NLP modules” such as the “NLP Manager” that are “the parser of the system.” EX1004, 8:27-35, 28:44-48. These modules perform natural language processing that “involves the lexical, syntactic (grammatical), and domain semantic analysis of the user input” corresponding to the “*first natural language utterance*,” “using both statistical observations of the various surface forms and a deeper interpretation of the relationships and dependencies among words, phrases, and concepts,” which a POSITA would have understood corresponds to “*parsing . . . the first natural language utterance.*” *Id.*

183. In addition, Kennewick also discloses this limitation. For example, Kennewick discloses that the speech processing unit (e.g., “the one or more processors”) includes “a parser for *parsing* the input” corresponding to the claimed “*first natural language utterance*.” EX1006, [0013]; *see also id.*, [0028] (“parsing and interpreting the question or query”). A POSITA would have understood that in the combined system discussed above, the one or more processors of the system may parse the first natural language in the same way as taught by Kennewick. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [18.b] obvious.

[18.c] “interpreting, by the one or more processors, the first natural language utterance; and”

184. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. Aretoulaki discloses that the system’s processing modules (e.g., the “*one or more processors*”) includes an “NLP Manager,” “Dialogue Manager,” and “Knowledge Manager.” EX1004, 8:4-7. As I discussed above for claim 5, these components correspond to “*a conversational language processor configured to interpret a natural language utterance*,” which a POSITA would have understood includes the claimed “*first natural language utterance*.” For example, the NLP Manager receives the “input text string” corresponding to the first natural language utterance, and outputs “frames with embedded structures holding the recognized words from the user input, their syntactic and semantic function and possible

semantic relationship among them.” *Id.*, 8:9-16, 8:40-44. These frames are an initial interpretation of the natural language utterance. *See also id.*, 8:29-35 (“Natural Language Processing (NLP) involves the lexical, syntactic (grammatical), and domain semantic analysis of the user input using both statistical observations of the various surface forms and a deeper interpretation of the relationships and dependencies among words, phrases, and concepts.”). In addition, the Dialogue Manager may perform further interpretation by augmenting the initial interpretation through discourse analysis using “user dialogue acts.” *Id.*, 12:15-35; *see also id.*, 9:39-44 (“Dialogue Manager 120 employs a series of models that help it interpret the user input and decide on the next system action.”). The Dialogue Manager and Knowledge Manager may select between “different interpretations . . . representing the various alternative meanings in the case of ambiguities,” using “[i]nformation about the context of the exchange . . . kept in the discourse history, and about the salient domain concepts in the knowledge base.” *Id.*, 15:16-30. Thus, Aretoulaki discloses “*interpreting, by the one or more processors, the first natural language utterance.*”

185. In addition, Kennewick also discloses this limitation. For example, Kennewick discloses that the parser included as part of its main speech processing unit (e.g., “*the one or more processors*”) performs “parsing and *interpreting* [of] the [user’s] question or query” corresponding to the claimed “*first natural language*

utterance.” EX1006, [0028]. Thus, Kennewick discloses “*interpreting, by the one or more processors, the first natural language utterance.*” A POSITA would have understood that in the combined system discussed above, the one or more processors of the system may interpret the first natural language in the same way as taught by Kennewick. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [18.c] obvious.

[18.d] “invoking, by the one or more processors, an applicable agent from a library of agents.”

186. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. Aretoulaki discloses that the system’s processing modules (e.g., the “*one or more processors*”) includes a “Knowledge Manager” that receives the interpretation of the first natural language utterance containing “the salient domain concepts and their values appearing in the user input,” and “dynamically accesses the application database” of an applicable application “to retrieve the best matches to the user’s requirements as specified up to that point.” EX1004, 15:44-49; *see also id.*, 8:50-52 (“Knowledge Manager 115 maintains and manipulates information on . . . the specific domain and application under consideration”). However, to the extent that Aretoulaki does not disclose the accessed application is an “*applicable agent invoked from a library of agents,*” the combination with Kennewick renders this obvious. Kennewick discloses invoking, based on the first natural language utterance, “the required agent” to process and respond to the user’s request, which

is selected from a library of agents 106 maintained by the speech processing unit. EX1006, [0028], [0032], [0127]-[0128]. For example, FIG. 6 of Kennewick illustrates an “agent architecture” that includes a collection or library of domain agents managed by an “agent manager.” See EX1006, [0126]-[0127], FIG. 6. A POSITA would have understood that the “required agent” disclosed by Kennewick is “*an applicable agent from a library of agents.*”

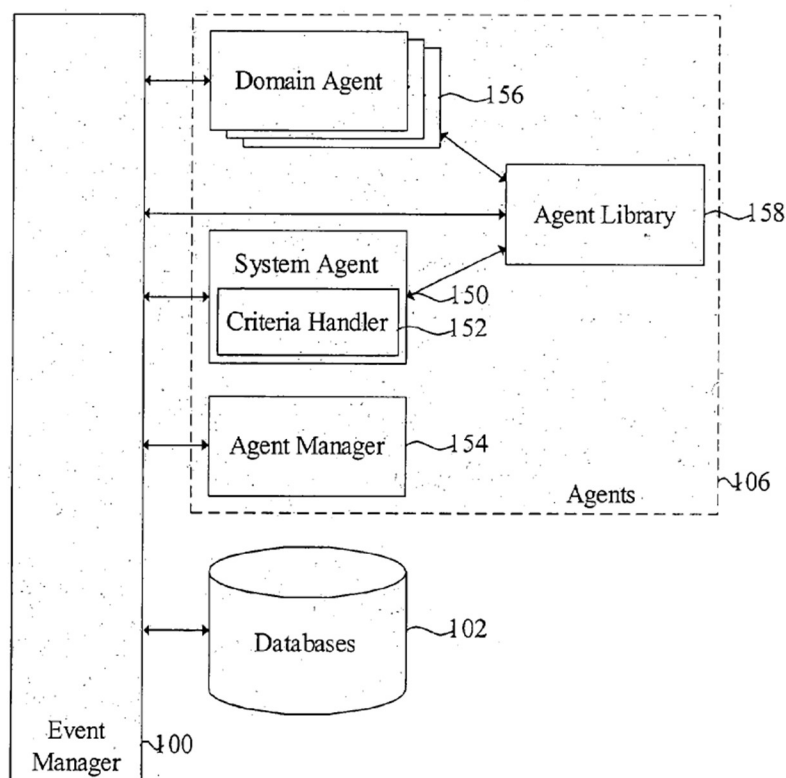


Figure 6. Agent Architecture

187. As I discussed above, a POSITA would have found it obvious in the combined system to modify Aretoulaki’s Knowledge Manager to utilize an agent architecture having domain agents configured to perform operations relating to

specific domains and applications as taught by Kennewick. In addition, as discussed above, a POSITA would have found it obvious that a collection of multiple domain agents as disclosed in Kennewick are a “*library of agents*” managed by Kennewick’s agent architecture. A POSITA would have been motivated to make this modification to enable Aretoulaki’s Knowledge Manager to more efficiently “maintain[] and manipulate[] information on . . . specific domain[s] and application[s] under consideration,” and to enable the system to be more easily extended to new domains and applications, e.g., by adding or updating the library’s domain agents. See EX1004, 8:50-52; EX1006, [0126] (“an agent may include everything needed to extend or modify the functionality of the system in a current or new domain”). For example, Aretoulaki discloses that received user inputs can relate to a variety of different domains and applications, and it would have been obvious to modify Aretoulaki based on the teachings of Kennewick to organize the domain-specific behavior and information managed by the Knowledge Manager into a library of different agents, and “*invoking . . . an applicable agent from [the] library of agents*” based on “the specific domain or application under consideration.” See EX1006, [0017] (“domain specific behavior and information are organized into agents”); EX1004, 8:50-52 (“Knowledge Manager 115 maintains and manipulates information on . . . the specific domain and application under consideration”). A POSITA would have had a reasonable expectation of success because the

combination is a simple substitution of one known technique (managing information for different domains through a Knowledge Manager) for another (managing information for different domains through respective domain agents) to obtain predictable results (managing information for different domains through a Knowledge Manager using respective domain agents). The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [18.d] obvious.

12. Claim 19

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

188. Aretoulaki in combination with Ramer and Kennewick renders the preamble obvious. As discussed above, Aretoulaki in combination with Ramer renders the method of claim 1 obvious.

[19.a] “invoking an agent to process the natural language utterance; and”

189. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. As discussed above for limitations [11.a]-[11.b] and [18.d], a POSITA would have found it obvious to combine Aretoulaki, Ramer, and Kennewick such that Aretoulaki’s Knowledge Manager invokes an agent corresponding to a determined domain or application to process the natural language utterance as taught by Kennewick. *See* EX1006, [0028] (describing “determining the domain . . . and context,” and then “invoking the proper resources, including agents”). Kennewick further teaches that the invoked agent “process[es] the commands or questions” corresponding to the natural language utterance. *Id.*,

[0123]. Thus, in the combined system, Aretoulaki's Knowledge Manager *"invoke[es] an agent to process the natural language utterance"* as taught by Kennewick. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [19.a] obvious.

[19.b] "creating a query to a local database or through a network interface to a remote database."

190. Aretoulaki in combination with Ramer and Kennewick renders this limitation obvious. For example, Aretoulaki discloses that the Knowledge Manager receives "the salient domain concepts and their values appearing in the user input (product/service description 255[...])" as part of a "domain semantic representation," and "dynamically accesses the application database 125 . . . to retrieve the best matches to the user's requirements as specified up to that point." EX1004, 15:1-9, 15:44-49. A POSITA would have found obvious that the Knowledge Manager accessing and retrieving information from a database would involve *"creating a query"* to the application database, which may be a *"local database"* that the Knowledge Manager "has direct access to." *See id.*, 14:38-40, 24:12-13 ("database query").

191. In addition, this limitation is further rendered obvious by the combined system discussed above, in which Aretoulaki's Knowledge Manager is modified to invoke an agent to *"process the natural language utterance"* as taught by Kennewick. For example, Kennewick discloses that the invoked agent "process[es]

the commands or questions” corresponding to the natural language utterance by “create[ing] queries to local databases 102 or through [a] network interface 116 to data sources on the Internet 42 or other networks” (e.g., a “*remote database*”). EX1006, [0123]. A POSITA would have understood that in the combined system, the invoked agent would “*creat[e] a query to a local database or through a network interface to a remote database,*” which in the combination may correspond to Aretoulaki’s “application database,” in the same way as taught in Kennewick. The combination of Aretoulaki, Ramer, and Kennewick therefore renders limitation [19.b] obvious.

13. Claim 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.”

192. Aretoulaki in combination with Ramer and Kennewick renders claim 20 obvious. For example, Aretoulaki discloses maintaining a “user model” containing “data about the general preferences and assumptions of individual users,” which, as I discussed above, corresponds to the claimed “*user-specific profile*” and/or “*preferences.*” EX1004, 9:47-49. However, to the extent that Aretoulaki does not explicitly disclose providing a user interface allowing the user to modify the user-specific profile, the combination with Kennewick renders it obvious. As discussed above, Kennewick discloses providing a “*user interface,*” e.g., a “natural language interactive speech interface,” allowing users to modify user-specific

“preferences” such as “the presentation formats for results.” EX1006, [0073]. In addition, Kennewick further discloses providing a user interface allowing the user to modify the user’s user-specific profile by “manually enter[ing]” user profile data including “user specific parameters.” *Id.*, [0142]-[0143]. Thus, Kennewick discloses *“providing a user interface allowing the user to modify the user-specific profile, questions, queries, data sources, preferences, or parameters”* through speech and/or manual entry.

193. A POSITA would have found it obvious to modify Aretoulaki to incorporate Kennewick’s user interface allowing the user to modify the user-specific profile and/or preferences into Aretoulaki’s system. For example, Aretoulaki discloses that the user model may include the user’s “presentation preferences.” EX1004, 9:49-52. A POSITA would have been motivated to incorporate Kennewick’s teachings of allowing users to modify their presentation preferences, e.g., “the presentation formats for results,” through a natural language interactive speech interface or other type of user interface. EX1006, [0073]. For example, in the combined system, the user may be able to use Aretoulaki’s existing natural language speech interface to modify their presentation preferences in the same manner as taught by Kennewick, and/or use Aretoutaki’s text-based interface to “manually enter[.]” preference information as taught by Kennewick. EX1004, Abstract; EX1006, [0143]. A POSITA would have been motivated to make this

combination to provide an obvious way for the user of Aretoulaki to provide and/or modify their profile and/or preferences, ensuring that the “user model” includes the user’s latest “preferences,” thus enabling the system to better meet its users’ needs. A POSITA would have found such a modification to be obvious, as it simply involves applying a known technique (e.g., the user being able to modify their user profile through a user interface) to a similar system (e.g., Aretoulaki’s system that maintains user profile information) in the same way to yield predictable results (e.g., the user being able to modify their maintained user profile, e.g., preferences, through the interface). The combination of Aretoulaki, Ramer, and Kennewick therefore renders claim 20 obvious.

14. Claim 28

194. Claim 28 is substantively identical to claim 6 and is obvious for the same reasons.

15. Claim 29

195. Claim 29 is substantively identical to claim 7 and is obvious for the same reasons.

16. Claim 33

196. Claim 33 is substantively identical to claim 11 and is obvious for the same reasons.

17. Claim 34

197. Claim 34 is substantively identical to claim 12 and is obvious for the same reasons.

18. Claim 35

198. Claim 35 is substantively identical to claim 13 and is obvious for the same reasons.

19. Claim 36

199. Claim 36 is substantively identical to claim 14 and is obvious for the same reasons.

20. Claim 37

200. Claim 37 is substantively identical to claim 15 and is obvious for the same reasons.

21. Claim 38

201. Claim 38 is substantively identical to claim 16 and is obvious for the same reasons.

22. Claim 39

202. Claim 39 is substantively identical to claim 17 and is obvious for the same reasons.

23. Claim 40

203. Claim 40 is substantively identical to claim 18 and is obvious for the same reasons.

24. Claim 41

204. Claim 41 is substantively identical to claim 19 and is obvious for the same reasons.

25. Claim 42

205. Claim 42 is substantively identical to claim 20 and is obvious for the same reasons.

D. Ground 4: Kennewick in combination with Hao render obvious claims 1-2, 4-7, 9-24, 26-29, and 31-44

1. General Motivations to Combine Kennewick and Hao

206. In my opinion, a POSITA would have been motivated to combine Kennewick and Hao. Both references relate to systems and methods for natural language processing and enabling users to conduct e-commerce by purchasing products or services by using their voice. *See, e.g.*, EX1006, Abstract (“a complete speech-based information query, retrieval, presentation and local or remote command environment”), [0018] (“natural language telematics speech interface” utilizes “[d]omain specific agents [that] include the behavior and data required for a specific area of functionality,” such as “travel services, entertainment scheduling . . . fast food ordering agent [that] may use the services of a restaurant ordering agent and payment agent”), [0060], [0066] (“interactive speech interface” can be employed to provide “a variety of services,” including “[l]ocation based marketing programs wherein occupants of the vehicle . . . can query the system for offers and

promotions for goods and services,” and “the system can provide interactive offers and promotions for goods and services”), [0067] (“Remote ordering and payment for goods and services. The system can interactively present the menu or product catalog using the list and table presentation capabilities of the invention. The system facilitates remote ordering by using location information, customer preferences, customer order history, etc.”), [0068] (“Travel services” that include “directories of travel and entertainment services, or reservations for entertainment restaurants, hotels and other accommodation.”); EX1007, Abstract (“an e-commerce shopping method that includes a step of receiving shopper free form input representing a shopping request” from which “semantics . . . relating to at least one item available for sale” are extracted), [0009] (“a solution designed to enhance e-commerce shopping so that shoppers can interact with an automated system using natural language input” that is “verbal . . . and can be conveyed during a real-time communication session”), [0011] (shopping cart “used to present items specified by the shopper for shopper purchase, wherein the presented items are determined from input provided via the natural language input element”), [0040] (“a shopper specific shopping cart can be populated with items ascertained from the shopping request,” where “the shopper can be presented with the shopping cart and can be permitted to utilize one or more of the shopping cart options” such as finalizing “an e-commerce transaction”).

207. Additionally, both Kennewick and Hao share similar architectures for an interactive speech system. Both disclose an input device that captures a user's voice input. *See, e.g.*, EX1006, [0121] (“speech unit 128 includes one or more microphones . . . to receive the utterances of the user”), claim 1 (“a speech unit that receives the user generated natural language speech utterance”); EX1007, [0021] (“client device 111 hosting a commerce interface 112”), claim 8 (received “[shopper] free form input is speech input in a natural language format”). Both disclose speech recognition components that perform recognition of the words and phrases of the user's voice input. *See, e.g.*, EX1006, [0123], [0155] (“[t]he speech recognition engine 120 may recognize words and phrases, using for example, information in the dictionary and phrase tables 112”); EX1007, [0022] (“speech input can be processed by an automatic speech recognition (ASR) engine”), claim 15 (“a speech recognition engine configured to convert the shopper provided utterances into text”). Both disclose interpreting the recognized words or phrases of the user's input to select and present a product or service to the user's device. *See, e.g.*, EX1006, [0028] (“parsing and interpreting the question or query”), [0066]-[0067] (“provid[ing] interactive offers and promotions for goods and services” based on user queries), [0155] (parser receives recognized words and phrases “for interpretation”); EX1007, [0022] (“NLU engine 132”), [0032] (“extract semantics from the input that relates to at least one item available for sale”), claim 1

(“automatically adding at least one item to an electronic shopping cart based upon the extracted semantics”). Both disclose enabling the user to purchase presented goods or services through the interactive speech system. *See, e.g.*, EX1006, [0018] (describing “fast food ordering,” “travel services,” etc.), [0067] (“[t]he system facilitates remote ordering” and “payment”); EX1007, [0022] (“[t]he shopper 110 can . . . confirm a purchase via options provided by the commerce interface 112”), claim 1 (“providing an ability to a corresponding shopper to automatically purchase the items contained in the electronic shopping cart”).

208. Furthermore, Kennewick explains that to more accurately interpret and respond to user requests, the system may make “significant use of context, prior information, domain knowledge, and user specific profile data.” EX1006, [0007], [0011], [0118]. For example, Kennewick recognizes that “some questions” are ambiguous and “can only be interpreted in the context of . . . the user’s history of interests and preferences.” *Id.*, [0007]. Kennewick further discloses facilitating “[r]emote ordering and payment for goods and services” using user-specific information such as “customer preferences” and “customer order history.” *Id.*, [0067]. Similarly, Hao recognizes that a shopper’s “past transaction history can be . . . used to increase accuracy of future transactions,” and describes specific techniques for interpreting and resolving ambiguities in a shopper’s current request “by using shopper specific information maintained within a shopper

profile/preferences data store,” such as “items that have previously been purchased by [the] shopper.” EX1007, [0033], [0040]-[0041]. By resolving these ambiguities using the user’s “past transaction history,” the system determines appropriate items to be placed in an electronic shopping cart and made available for the shopper to purchase. *See id.*, [0039]-[0041], [0046].

209. A POSITA would have been motivated to incorporate Hao’s techniques for interpreting and resolving ambiguous user requests based on user-specific information such as “past transaction history” to select and present items for purchase to the user, to enhance and improve Kennewick’s use of “user specific profile data” such as “customer preferences” and “customer order history” to facilitate remote ordering of goods or services, even when the user’s requests are “incomplete, ambiguous or subjective.” *See* EX1007, [0041]; EX1006, [0002], [0007], [0011], [0067]. For example, a POSITA would have found it obvious to modify Kennewick, to the extent it does not describe specifically how to use “customer preferences” and “customer order history” to facilitate “[r]emote ordering and payment for goods and services,” to incorporate Hao’s techniques of leveraging the user’s “past transaction history . . . to increase the accuracy of future transactions” by resolving “ambiguities” in a current utterance using “historical shopping information.” *See* EX1006, [0067]; EX1007, [0027], [0033], [0041]. By resolving these ambiguities, the system can identify “one or more items” to present

to the user as part of a selected purchase opportunity, without having to solicit the user for additional information, thus “minimiz[ing] the interaction with the user required to deliver a response.” EX1007, [0039], [0045]-[0046]; EX1006, [0032]. This modification is a substitution of one known element (e.g., Hao’s specific teachings on ways to use “past transaction history” to disambiguate user requests and select items for purchase) for another (e.g., Kennewick’s more general disclosure of using “customer preferences” and “customer order history” to facilitate “[r]emote ordering”) to yield predictable results (e.g., Kennewick’s system using “customer preferences” and “customer order history” to facilitate “[r]emote ordering” by disambiguating user requests to select appropriate items for purchase as taught by Hao, thus “minimiz[ing] the interaction . . . required to deliver a response”). I provide additional motivations to combine the references below.

2. Claim 1

[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:”

210. To the extent that the preamble of claim 1 is limiting, Kennewick in combination with Hao renders it obvious. First, Kennewick discloses a “***method for processing natural language utterances that include requests.***” For example, Kennewick discloses methods “for processing of commands through a speech

interface” through which users can “ask natural language speech questions or give natural language commands in a wide range of domains” corresponding to different areas of functionality. EX1006, [0002], [0018]. A “main speech processing unit” receives the user input (e.g., “natural language queries, commands and/or other utterances”) from the speech interface “and processes the input” to “retrieve[] information responsive to the query” and/or “take[] action consistent with the command.” *Id.*, [0014]; *see also id.*, FIGs. 4-5 (illustrating a “main speech processing unit 98” or “main unit 98”). The “natural language speech questions” and “natural language speech commands” meet the claimed “*natural language utterances that include requests.*”

211. Second, Kennewick discloses “*selecting and presenting purchase opportunities based thereon*” (i.e., based on the processed natural language utterances). For example, Kennewick’s system offers services like “[r]emote ordering and payment for goods and services” and “[l]ocation based marketing programs.” EX1006, [0066]-[0067]. Through these services, users can “query the system for offers and promotions for goods and services.” *Id.*, [0066]. The system will then “apply other available information” such as “users profile, history and location” to select and “provide interactive offers and promotions for goods and services along the route.” *Id.* The system may also “interactively present [a] menu or product catalog” to the user, and “facilitate[] remote ordering by using location

information, customer preferences, customer order history, etc.” *Id.*, [0067]. These presented offers, promotions, and catalogs are “*purchase opportunities*” that Kennewick’s system selects and presents to the user based on the user’s natural language queries and/or requests. For example, Kennewick explains that the system selects and presents “interactive offers and promotions for goods and services” based on the user’s “query . . . for offers and promotions for goods and services,” and/or selects and presents goods for “[r]emote ordering and payment” based on user requests received through “the interactive speech interface of the system.” *Id.*, [0060], [0066]-[0067].

212. Kennewick further discloses that the method is “*implemented by one or more physical processors.*” For example, Kennewick discloses that the user interacts with the speech interface through a “speech unit” that communicates with a “main speech processing unit” that processes the received utterances. EX1006, [0121]-[0123]. These two units may be “embedded into” a “computer” that includes “one or more processing units.” *Id.*, [0111], [0115] (“the main speech processing unit 98 can be integrated with the one or more processing units 84”). For example, FIG. 4 of Kennewick (reproduced below), shows the speech unit 128 and the main speech processing unit 98 embedded into a “fixed computer.” *Id.*, FIG. 4; *see also id.*, [0122] (“[t]he speech unit can be integrated with the main [speech processing] unit”).

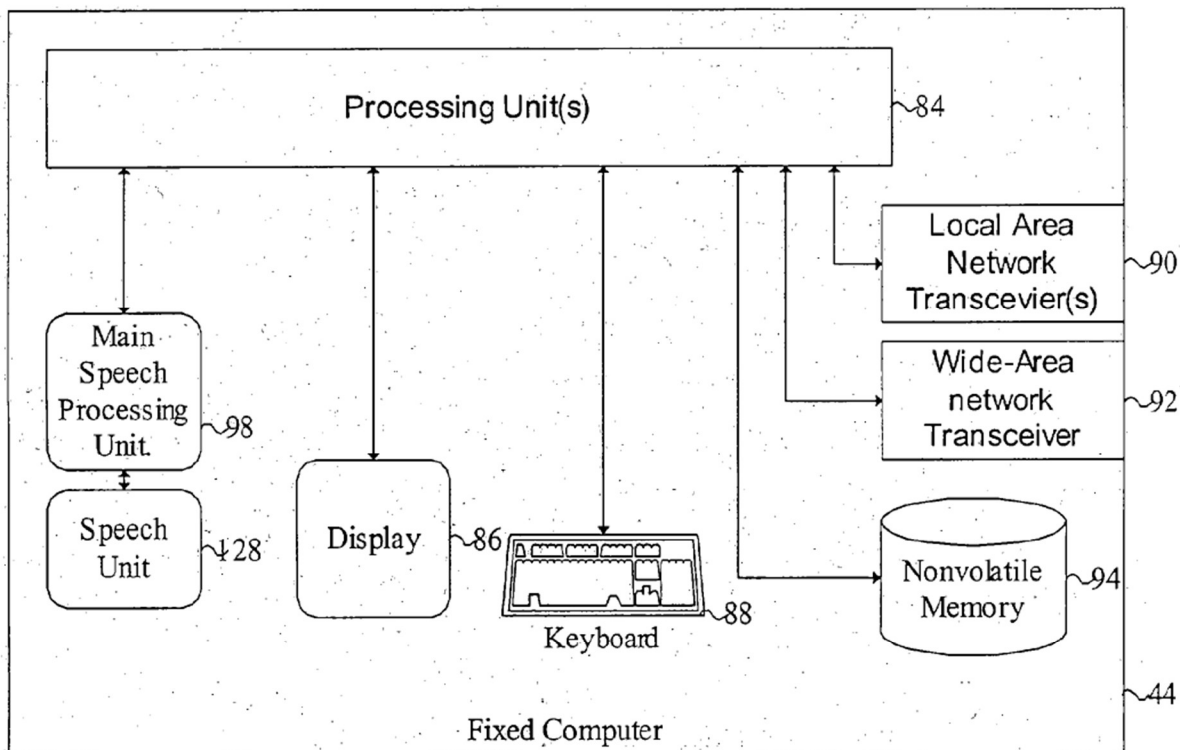


Figure 4. Fixed Computer Block Diagram.

EX1006, FIG. 4.

213. The “fixed computer” shown in the FIG. 4 includes “*one or more physical processors*,” including main speech processing unit 98, speech unit 28, and processing unit(s) 84. Thus, Kennewick discloses the method being implemented by one or more physical processors, because the method is implemented by the main speech processing unit which is a physical processor that also interfaces with or is integrated with one or more additional physical processors (“one or more processing units 84”).

214. Kennewick further discloses that the one or more physical processors are “*programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method.*” For example, Kennewick’s “*one or more physical processors,*” including the speech unit 128, main speech processing unit 98, and/or one or more processing units 84, contain various software modules like “a speech recognition module,” “a parser,” “a text to speech engine” as well as “agents” that are “packages of executable code.” EX1006, [0013], [0017]. These software modules and packages of executable code correspond to the claimed “computer program instructions” that, when executed by the one or more processors of Kennewick’s main speech processing unit, cause the one or more processors to perform various method steps. Kennewick therefore renders limitation [1.pre] obvious.

215. In addition, Hao also discloses a “*method for processing natural language utterances that include requests and selecting and presenting purchase opportunities based thereon.*” Hao discloses an “e-commerce shopping method” that includes “receiving shopper free form input representing a shopping request.” EX1007, Abstract. Hao’s e-commerce interface receives shopping requests that the user conveys verbally “in a natural language, such as English,” which meets the claimed “*natural language utterances that include requests.*” *Id.*, [0009], claim 8 (“the free form input is speech input in a natural language format”). Hao further

discloses processing the requests using “an automatic speech recognition (ASR) engine 134 and then further . . . by a natural language understanding (NLU) engine.” *Id.*, [0022]. Hao thus discloses a “***method for processing natural language utterances that include requests.***”

216. Hao further discloses “***selecting and presenting purchase opportunities based thereon***” (i.e., based on the processed natural language utterances). Hao’s e-commerce interface processes a natural language utterance request by “extract[ing] semantics from” the utterance “relating to at least one item available for sale,” and “automatically add[ing]” one or more items selected “based upon the extracted semantics” to “an electronic shopping cart.” *Id.*, [0012]. The e-commerce interface presents the electronic shopping cart containing the selected items to the user. *Id.*, [0040]. This meets the claimed “***purchase opportunities***” because the user is provided the opportunity to “finalize . . . an e-commerce transaction involving a purchase of the items presented in the shopping cart.” *Id.*, [0011], [0040]. Hao thus discloses “***selecting and presenting purchase opportunities based***” on the processed natural language utterances because it discloses processing verbal natural language requests to select and present purchase opportunities.

217. Hao also discloses that its method is “***implemented by one or more physical processors.***” For example, Hao’s method is implemented by a system that

includes an “input processing system” (or “input processor”) that receives and semantically interprets free form shopper input, and an “e-commerce server” that “conduct[s] an e-commerce transaction based upon this input.” EX1007, [0010], [0032]. These components correspond to “*one or more physical processors*” that implement the method by processing “free form shopper input” corresponding to natural language utterances that include requests, e.g., a “shopping request,” to select and present “at least one item available for sale” as part of a purchase opportunity based thereon. *Id.*, [0010], [0012]. For example, FIG. 1 (reproduced below) illustrates the “Input Processing System 130” and “E-commerce Server 120” that receive and process the user’s “Free Form Shopping communications 152” to select and present purchase opportunities to the user.

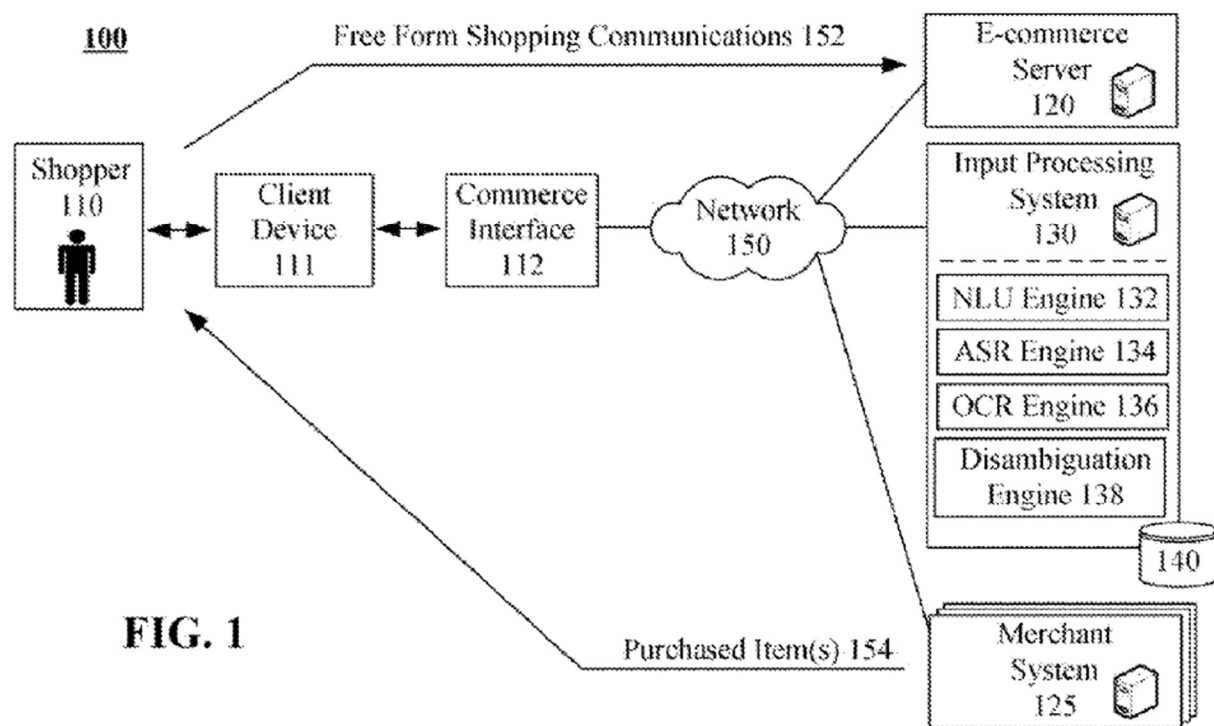


FIG. 1

Id., FIG. 1.

218. Hao further discloses that the one or more physical processors are ***“programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method.”*** For example, Hao discloses that the components of the system such as the input processing system and e-commerce server “may be realized in hardware, software, or a combination of hardware and software,” such as a “general purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein.” EX1007, [0054]. The “computer program” meets the claimed ***“computer program instructions,”*** which is

executed by the one or more physical processors of the system to perform the various method steps. *Id.* Hao therefore renders limitation [1.pre] obvious.

219. Thus, Kennewick and Hao each render obvious the preamble. In addition, a POSITA would have found it obvious to combine Kennewick with Hao, as each reference discloses similar speech systems that perform natural language processing and enable users to conduct e-commerce by purchasing products or services through voice. For example, Kennewick discloses processing received utterances to select and present purchase opportunities (e.g., “interactive offers and promotions”) and facilitate the “ordering and payment” of presented goods or services. EX1006, [0066]-[0067]. To the extent that Kennewick does not explicitly describe how to select these purchase opportunities, a POSITA would have found it obvious to look to the teachings of Hao for ways to do so, and would have been motivated to combine Kennewick, which leaves open its specific method for selecting “interactive offers and promotions,” with Hao’s disclosed methods for providing those purchase opportunities. A POSITA would have found such a combination 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 to yield predictable results, so that Kennewick’s system, upon determining that the user’s request relates to remote ordering/purchasing, selects and presents a purchase opportunity based on

the techniques taught in Hao. Kennewick in combination with Hao therefore renders limitation [1.pre] obvious.

[1.a]: “providing, by the one or more processors, a natural language utterance as an input to a speech recognition engine;”

220. Kennewick in combination with Hao renders this limitation obvious. As explained for limitation [1.pre] above, the “*one or more processors*” in Kennewick includes the main speech processing unit. The main speech processing unit “receives spoken natural language queries, commands and/or other utterances from a user” from a “speech coder” 138 of the “speech unit” 128 as a “speech signal,” and provides it as an input to a “*speech recognition engine*” 120. EX1006, [0014], [0149], [0155]. For example, FIG. 5 below shows the main speech processing unit 98 (“main unit 98”) receiving a speech signal from the speech unit 128 at the speech coder 122 and providing it to the speech recognition engine 120.

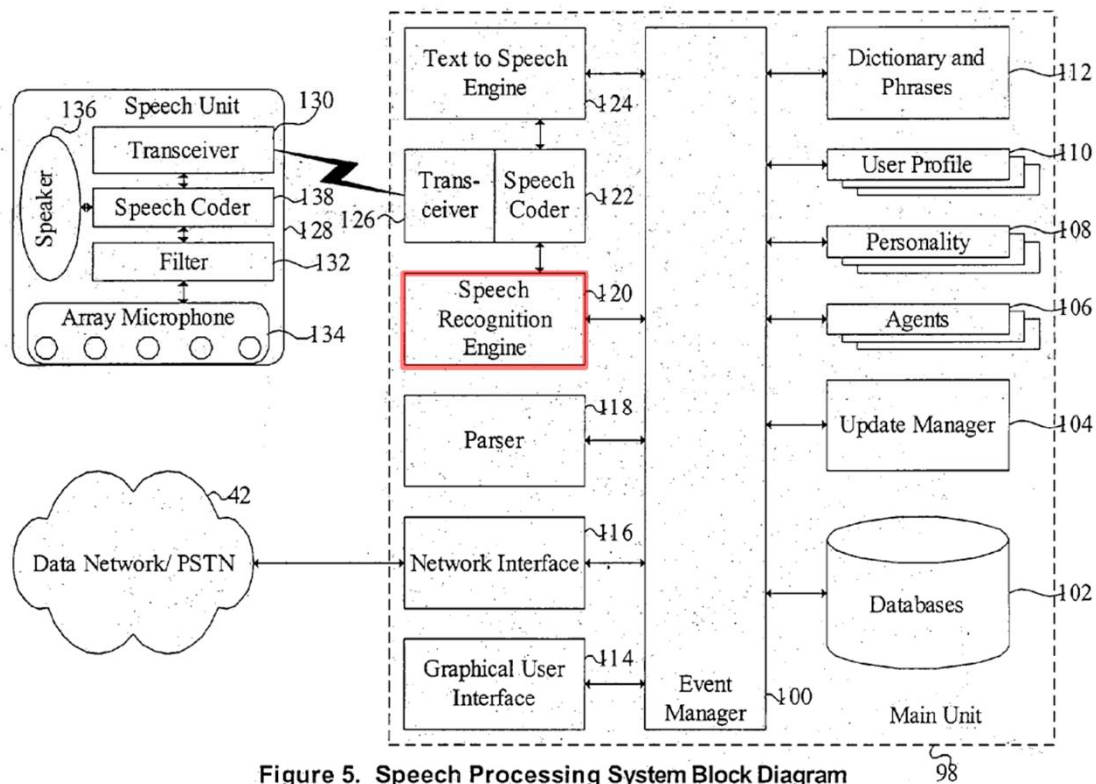


Figure 5. Speech Processing System Block Diagram

Id., FIG. 5. Kennewick in combination with Hao therefore renders limitation [1.a] obvious.

[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;”

221. Kennewick in combination with Hao renders this limitation obvious. As explained for limitation [1.pre] above, the “*one or more processors*” in Kennewick include the main speech processing unit. The main speech processing unit includes a speech recognition engine (e.g., speech recognition engine 120). EX1006, [0123], FIG. 5. The speech recognition engine processes the received natural language utterance “using data in the dictionary and phrases modules 112”

to generate and output “recognized words and phrases” corresponding to the claimed *“words or phrases, recognized from the natural language utterance.”* EX1006, [0123], [0155] (“The speech recognition engine 120 may recognize words and phrases”). The speech recognition engine sends the words and phrases to event manager 100, a component of the main speech processing unit. *Id.*, [0119]. That event manager then sends the words and phrases to other components of the main speech processing unit, like “the parser 118 for interpretation.” *Id.*, [0155], [0119] (“The event manager 100 may mediate interactions between other components of the” main speech processing unit); *see also id.*, claim 2 (“event manager coordinating interaction between components of the” main speech processing unit). Thus, Kennewick discloses that the main speech processing unit (the *“one or more processors”*) receives the words or phrases recognized from the natural language utterance as an output of the speech recognition, which it passes to other components such as the parser 118. *Id.* Kennewick in combination with Hao therefore renders limitation [1.b] obvious.

[1.c] “determining, by the one or more processors, a context for the natural language utterance based on the recognized words or phrases;”

222. Kennewick in combination with Hao renders this limitation obvious. As explained for limitation [1.pre] above, the *“one or more processors”* in Kennewick includes the main speech processing unit. Kennewick discloses that the main speech processing unit determines *“a context for the natural language*

utterance” by “determin[ing] the most[] likely context or domain for a user’s question or command, for example, by using a real-time scoring system or other techniques.” EX1006, [0032]. For example, the main speech processing unit includes a “parser 118” which receives “tokens” corresponding to the recognized words or phrases of the natural language utterance. EX1006, [0160]. The speech recognition engine 120 within the main processing unit 98 generates the tokens by recognizing the words and phrases of the natural language utterance and identifying certain keywords as “tokens.” *See id.* The parser 118 “examin[es] the tokens for . . . context.” *Id.*, [0160]. It “determine[s] a context for [the] utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents.” *Id.*, [0160]. For example, the parser may determine that based on the natural language utterance containing the keyword “temperature,” that the utterance is associated with a weather context or a measurement context, and may “use a scoring system to determine the most[] likely context or domain for a user’s question or command” based on additional factors such as “the user profile 110, the domain agent’s 156 knowledge and previous context.” *Id.*, [0160]-[0161]. The determined context is used to “determine the domain and thereby, the domain agent 156, if any, to be invoked” to further process and respond to the utterance. *Id.*; *see also id.*, [0032] (the system “determine[s] the most[] likely context or domain for a user’s question or command” to “invoke the

correct agent” for returning “a formatted response”), [0127] (parser “invoke[s] the required agent” based “on the keywords and command structure” of the user’s utterance).

223. Kennewick’s disclosure of context and how it is determined is consistent with the disclosure of the ’758 patent, which describes determining the context for an utterance “by having domain agents 135 competitively generate a context-based interpretation of the utterance,” and invoking a domain agent associated with “a highest scoring interpretation” to further process and respond to the utterance. EX1001, 5:4-14, claim 6. The ’758 patent further explains that generating the “context-based interpretations” of an utterance is described in U.S. Patent No. 7,640,160 (“Di Cristo”), which it incorporates by reference. *Id.*, 5:14-25. Di Cristo explains that a parser “determines a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, and dialog history,” and that the “context of a question or command determines the domain and thereby, the domain agent 156, if any, to be evoked,” which is identical to Kennewick’s disclosure of how the parser determines a context. EX1014, 20:63-21:2; EX1006, [0160].

224. Thus, Kennewick discloses “*determining . . . a context for the natural language utterance based on the recognized words or phrases*” by examining and performing “keyword matching” on the tokens corresponding to the recognized

words or phrases (e.g., based on received “keyword and associated prior probability or fuzzy possibility updates”) to determine the context of the utterance. *Id.*, [0160].

Kennewick in combination with Hao therefore renders limitation [1.c] obvious.

[1.d] “selecting, by the one or more processors, a purchase opportunity based on the determined context;”

225. Kennewick in combination with Hao renders this limitation obvious. As discussed above, Kennewick’s “main speech processing unit” meeting the claimed “*one or more processors*” includes a parser configured to “*determin[e] . . . a context for the natural language utterance based on the recognized words or phrases*” of the utterance. *Id.*, [0102]. Based on the determined context, the parser “determine[s] the domain and thereby, the domain agent 156, if any, to be invoked” to further process and respond to the utterance. EX1006, [0160]; *see also id.*, [0127] (parser “invoke[s] the required agent” based “on the keywords and command structure” of the user’s utterance). Each invoked domain agent corresponds to a service for providing a respective “area of functionality,” which include e-commerce services through which the user may perform “[r]emote ordering and payment for goods and services.” *Id.*, [0018], [0060]-[0069]. For example, the user may purchase fast food through a “fast food ordering agent,” travel services through “a travel services agent,” and/or other types of goods or services through a “[r]emote ordering and payment” service. *Id.* The user “quer[ies] the system for offers and promotions for goods and services” through these services, and the system responds

by “provid[ing] interactive offers and promotions for goods and services” based on the user’s request. *Id.*, [0066]-[0067]. The offers and promotions for goods and services the system provides correspond to a selected “***purchase opportunity***” as the user is provided the opportunity to purchase the selected good or service. *See also id.*, [0067] (“The system facilitates remote ordering” and “payment”). In addition, the offers and promotions for goods and services the system provides (corresponding to the “***purchase opportunity***”) are selected “***based on the determined context***,” e.g., selected based on the determined context corresponding to one of Kennewick’s e-commerce services, such as the remote ordering and/or location-based marketing domain. *See also id.*, [0018] (describing separate services for “fast food ordering,” “travel services,” etc.), [0060]-[0069] (describing different applications supported by the “natural-language interactive speech interface of the invention”).

226. In addition, to the extent that Kennewick does not explicitly disclose how its system selects a purchase opportunity based on the determined context, the combination with Hao renders it obvious. Hao discloses an “input processing system” that meets the claimed “***one or more processors***.” The input processing system processes a received “shopping request” to “identify a series of tokens . . . that have meaning in an e-commerce context.” EX1007, [0038], [0043]. By processing the tokens in the e-commerce context, the system determines “one or more shopping items . . . from the shopping request” and matches the “extract[ed]

semantics from the input” with “[a]vailable items . . . specified an inventory data store 223 of an e-commerce server 222.” *Id.*, [0032]-[0033], [0038]-[0040]. These may include “unambiguous items 440” for which item attributes associated with the analyzed tokens are clear, or “ambiguous items 445” which the system further processes using user-specific information such as the user’s “purchase history, preference, and profile information” to resolve the ambiguity. *See id.*, [0045]-[0046]. The input processing system loads the selected items “into a shopping cart” and presents the shopping cart to the user, which meet the claimed “***purchase opportunity***” as the user given the opportunity to interact with the shopping cart to “finalize [an] e-commerce transaction” to purchase the presented items. *Id.*, [0038]-[0040], [0045]. The input processing system selects the purchase opportunity “***based on the determined context***” of the natural language utterance being an e-commerce context, as it analyzes the meanings of words or phrase of the utterance in the e-commerce context. *See id.*, [0043].

227. In the combined system, it would have been obvious to apply Hao’s teaching of selecting available items based on extracted meanings of the user’s input in a given e-commerce context to improve and enhance Kennewick’s services for remote ordering of goods or services. For example, in the combined system, Kennewick’s main speech processing unit processes a received natural language utterance to determine a context for the natural language utterance, e.g., through

“applying prior probabilities or fuzzy possibilities to keyword matching” as discussed above, to establish which “domain” corresponding to a specific “area of functionality” should be invoked. EX1006, [0018], [0160]. Kennewick further discloses that each domain may be associated with an agent that “receive[s], process[es] and respond[s] to user questions, queries and commands” and manages “the behavior and data” “in [the] specific domain.” *Id.*, [0017]-[0018]. Hao teaches specific techniques for receiving, processing, and responding to user requests in an e-commerce context to select and present purchase opportunities to a user. *See, e.g.*, EX1007, [0032]-[0033], [0038]-[0040], [0045]-[0046]. A POSITA would have found it obvious to apply these teachings to one or more of Kennewick’s agents responsible for managing the domain-specific behavior of an e-commerce domain, such as a fast food ordering agent, remote ordering agent, etc., to select which “offers and promotions for goods and services” to present to the user in the e-commerce context of that domain. EX1006, [0066].

228. For example, in the combined system, responsive to the determined context being an “e-commerce context” corresponding to one of Kennewick’s domains, such as remote ordering, fast food ordering, or travel services, a POSITA would have found it obvious for the invoked agent to identify tokens of the utterance “hav[ing] meaning in [the determined] e-commerce context,” and compare extracted semantics “to inventory items available to the shopper 110 for purchase from at least

one merchant system” to select matching items to present to the user as a purchase opportunity as taught by Hao. EX1007, [0022], [0043]. This is consistent with Kennewick’s existing disclosure that invoked agents generally “formulat[e] one or more queries to one or more . . . data sources” to obtain “one or more results” for presentation to the user. EX1006, [0028], [0123]-[0124]. Thus, the combined system renders obvious selecting a “***purchase opportunity based on the determined context,***” by selecting a purchase opportunity based upon the determined context corresponding to one of Kennewick’s domains relating to e-commerce, based on the specific techniques disclosed by Hao.

229. Furthermore, Kennewick discloses that the system may invoke different contexts/domains through which different types of goods or services that can be purchased, such as “fast food” or “travel services,” with each domain having access to “domain specific information.” EX1006, [0018], [0029]. A POSITA would have found it obvious that in the combined system, the system would select the purchase opportunity based on “tokens” of the user’s natural language input having meaning in the determined context. For example, if the system determines the context of the natural language utterance to be “fast food,” a POSITA would have found it obvious for it to select a purchase opportunity relating to fast food, whereas if it determines that the context is “travel,” it would select a purchase opportunity relating travel products or services.

230. In my opinion, a POSITA would have been motivated to combine Kennewick with Hao, to integrate Hao's teachings of specific ways to select and present a purchase opportunity based on a received utterance associated with a given context with Kennewick's interactive speech interface. As discussed above, the references are analogous, both relating to natural language interfaces through which users can conduct e-commerce by purchasing products or services using natural language utterances. *See, e.g.*, EX1006, Abstract, [0018], [0060]-[0068]; EX1007, Abstract, [0009], [0011], [0040]. The combination is a combination of prior art elements disclosed in Kennewick and Hao according to known methods to yield predictable results. For example, Kennewick teaches receiving natural language utterances across multiple domains, determining a context for each utterance, which may include contexts for remote purchasing and e-commerce, and invoking an appropriate domain agent to process and respond to the utterance. *See* EX1006, [0028], [0118], [0160]. Hao discloses processing and responding to utterances "that have meaning in" a particular context, i.e. "an e-commerce context," by extracting semantic meanings from the utterance, and selecting items for purchase based on the extracted semantics. EX1007, [0022], [0043]. Hao further discloses specific techniques for detecting whether the user's request refers to "ambiguous items," and for automatically resolving detected ambiguities using the user's "purchase history, preference, and profile information" without needing to solicit additional

information from the user. *Id.*, [0039], [0045]-[0046]. A POSITA would have been motivated to make this combination as it provides a straightforward way for Kennewick's system to select a purchase opportunity for presentation to the user based on determining that the user's natural language utterance pertains to an e-commerce context, even if the user's request contains ambiguities. A POSITA would have had a reasonable expectation of success because the combination is a simple combination of prior art elements (establishing a context of an utterance 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 (e.g., Kennewick's system establishing a context of an utterance and then selecting a purchase opportunity based thereon). A POSITA would also have been motivated to make this combination, which applies a known technique (e.g., Hao's techniques for selecting a purchase opportunity and resolving ambiguities) to improve a similar system (e.g., Kennewick's interactive speech interface through which a user can perform remote ordering of goods or services) in the same way, to create a more robust interface that is able to automatically resolve ambiguities in the user's request to select and present a purchase opportunity. A POSITA would further have been motivated to make the modification as it aligns with Kennewick's teaching of using "user profile information . . . to minimize the interaction with the user required to

deliver a response.” EX1006, [0032]. The combination of Kennewick and Hao thus renders limitation [1.d] obvious.

[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;”

231. Kennewick in combination with Hao renders this limitation obvious. Kennewick’s main speech processing unit (e.g., “*the one or more processors*”) delivers the selected purchase opportunity via an electronic device in communication with the main speech processing unit. For example, and with reference to FIG. 5 below, Kennewick discloses that the main speech processing unit delivers a response to the user’s utterance by transmitting coded speech “from the main unit 98 . . . to a transceiver 130 on [a] speech unit 128,” where the speech is “decoded and decompressed by the speech coder 138 and output by the speaker 136” of the speech unit 128.

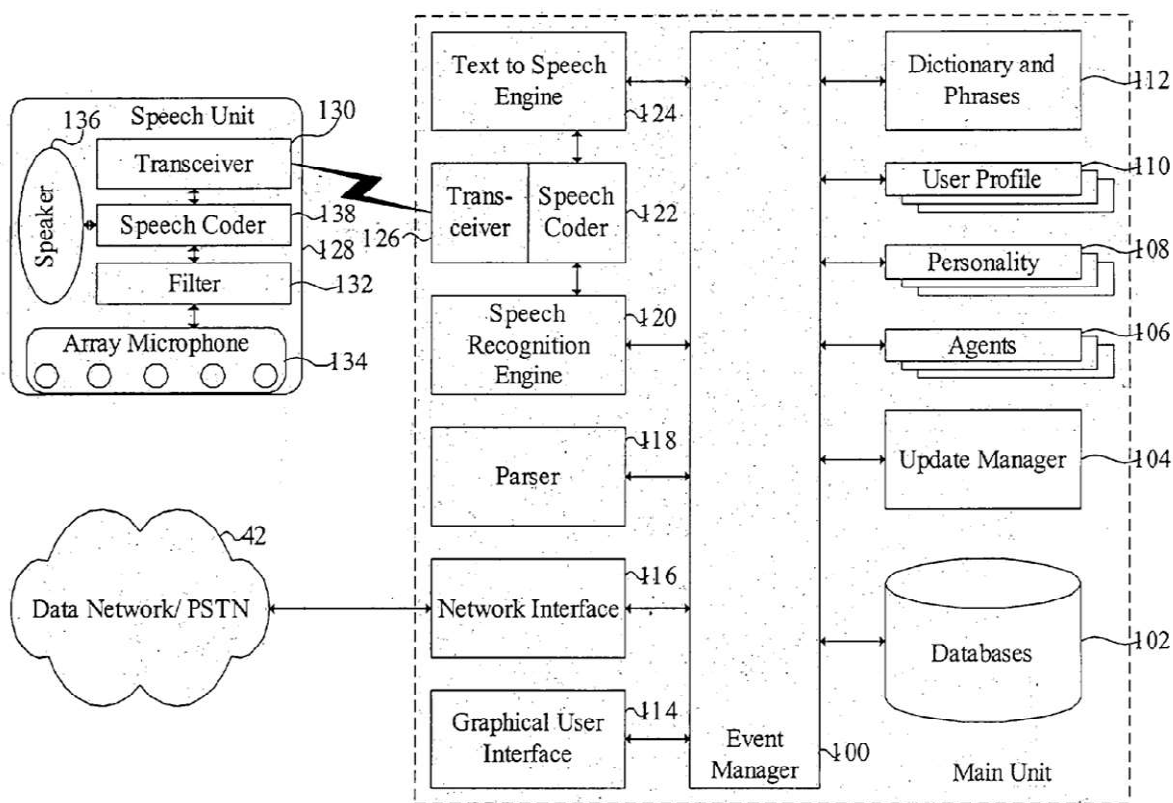


Figure 5. Speech Processing System Block Diagram

EX1006, [0121], [0124], FIG. 5. The main speech processing unit may also deliver the response to the user via a “graphical user interface 114” that “can be used as a substitute or complement to the speech interface” by displaying information via “a display 18” in a “graphical or tabular” manner “more easily digested by the user.” *Id.*, [0125]. As shown in FIG. 1 of Kennewick, the speech unit 128 and/or display 18 are “*electronic device[s] in communication with*” the main speech processing unit 98 (“*one or more processors*”).

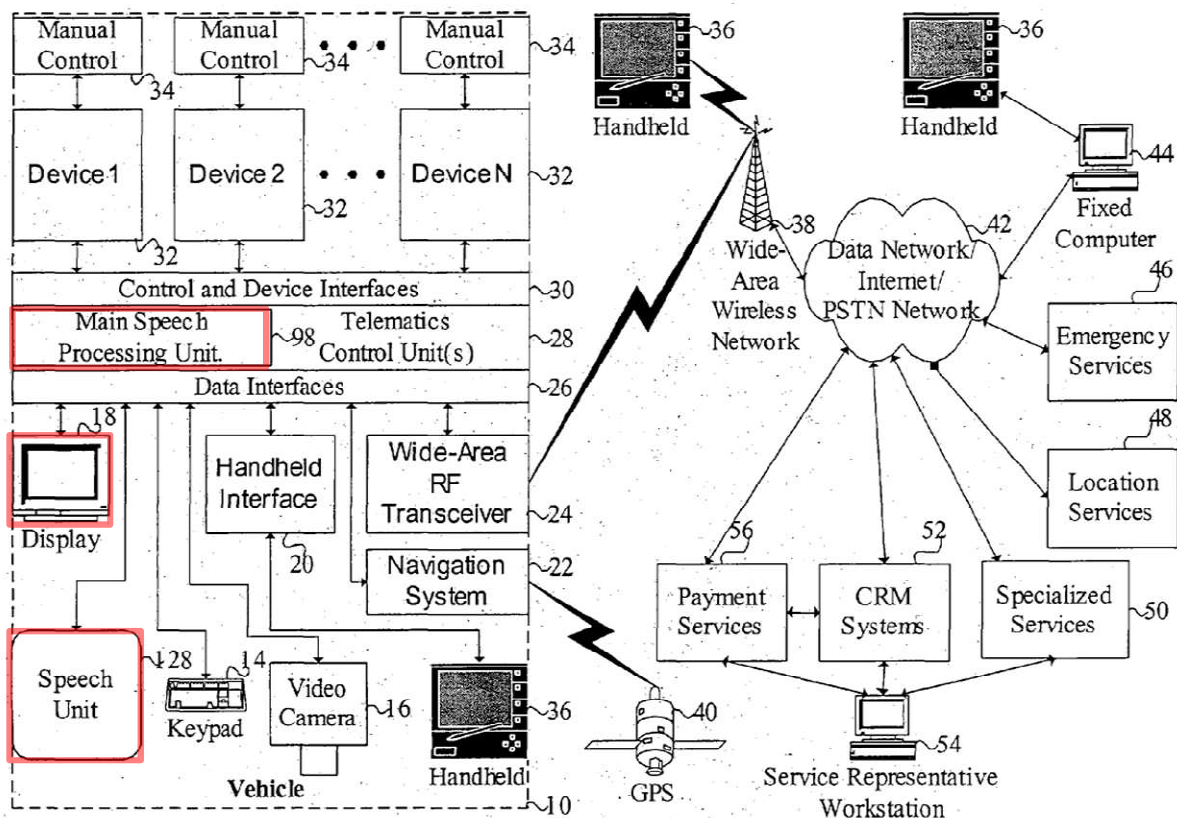


Figure 1. First Embodiment System Block Diagram

Id., FIG. 1.

232. The delivered response corresponds to the “*selected purchase opportunity*.” For example, as discussed above, Kennewick’s system provides services like “[l]ocation based marketing programs” and “[r]emote ordering and payment” through which the user can receive, as a response to their queries, purchase opportunities such as “promotional offers” and/or “product catalog[s].” *Id.*, [0066]-[0067]. These purchase opportunities are presented to the user via delivery to the “speech unit 128,” and/or to the graphical user interface of the display 18, e.g., using the “list and table presentation capabilities of the invention” to present selected

products or services as a list or table. *Id.*, [0066]-[0067]; *see also id.*, [0145] (“reports can be created using only voice commands and responses, commands and responses through the graphical user interface 114, or a combination of the two”).

233. In addition, the combination with Hao also renders this limitation obvious. For example, Hao discloses that after the system determines “one or more shopping items . . . from the shopping request,” the system populates a “shopper specific shopping cart . . . with items ascertained from the shopping request,” which as discussed above corresponds to “***the selected purchase opportunity***.” EX1007, [0038], [0040]. Hao’s system presents the shopping cart to the user through a “voice only interface” or through “a Web based interface accessible via a browser having a graphical user interface (GUI)” provided by a “client device” (corresponding to the claimed “***electronic device***”). *See id.*, claims 19-20. The user, when presented “with the shopping cart,” is “permitted to utilize one or more of the shopping cart options.” *Id.*, [0040].

234. A POSITA would have found it obvious for Kennewick’s main speech processing unit in the combined system to deliver an electronic shopping cart populated with one or more selected items as taught by Hao (corresponding to “***the selected purchase opportunity***”) to the user via the speech unit and/or the display as taught by Kennewick (corresponding to the “***electronic device in communication with the one or more processors***”). For example, as discussed above for limitation

[1.d], a POSITA would have found it obvious to modify Kennewick's main speech processing unit so that, upon determining that the natural language utterance relates to an e-commerce context, it selects items to populate an electronic shopping cart as taught by Hao to present to the user as a purchase opportunity. In this combination, a POSITA would have been found it obvious to deliver the electronic shopping cart to the user as a voice response and/or visually "using the list and table presentation capabilities" of Kennewick's system, as both references disclose responding to the user through a voice interface, a graphical interface, or some combination thereof. See EX1006, [0067], [0145]; EX1007, claims 19-20, [0021]. Kennewick in combination with Hao thus renders limitation [1.e] obvious.

[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;"

235. Kennewick in combination with Hao renders this limitation obvious. First, Kennewick discloses that the main speech processing unit (e.g., "***one or more processors***") tracks user interactions with the system to build "user profiles and histories" that include data such as the user's "history of questions asked, session histories," etc. EX1006, [0016], [0031]; *see also id.*, [0143] ("Personal profiles may include information specific to the individual, their interests, their special use of

terminology, the history of their interactions with the system, and domains of interest.”). The system “add[s] and update[s]” information of the tracked interactions “automatically as the user uses the system.” *Id.*, [0031].

236. A POSITA would have found it obvious that Kennewick’s tracking the history of the user’s interactions with the system may include “***tracking . . . an interaction pattern associated with the purchase opportunity delivered to the electronic device.***” For example, as discussed above, Kennewick discloses that its e-commerce-related services deliver a purchase opportunity to the electronic device. *See* EX1006, [0066]-[0067]. Kennewick further discloses tracking historical information about the user, such as the user’s “order history,” to provide “offers and promotions for goods and services” and facilitate “remote ordering” of goods and services. *Id.*, [0031], [0066]-[0067]. A POSITA would have understood that a user’s past purchase/order of goods or services corresponds to an “***interaction pattern associated with [a previously presented] purchase opportunity***” because a POSITA would have found it obvious that the user ordering a good or service associated with the previously presented purchase opportunity would include one or more interactions including the user making “***a subsequent request in which the electronic device,***” e.g., the speech unit, display, and/or keyboard through which the user interacts with the main speech processing unit, “***interacts with the [presented] purchase opportunity.***” *See* EX1006, [0112], [0115]-[0016] (users interact with the

main speech processing unit 98 “through the speech unit 128, [a] keyboard 88 or keypad, and[/or] a display 86 for text, graphics and video”), [0125] (user may interact with the “graphical user interface” via the “display 18, keypad 14, and[/or] pointing device”).

237. In addition, a POSITA would have understood that the subsequent request, i.e., the user’s request to purchase the good or service, results in the electronic device “*completing a transaction related to the purchase opportunity*” because Kennewick discloses “remote ordering” using a “secure payment wallet for the users” to pay for the purchase. *See* EX1006, [0067]. A POSITA would thus have found it obvious that Kennewick’s system tracks this interaction pattern associated with a purchase opportunity delivered to the user’s electronic device, comprising information of the user’s subsequent request to purchase a good or service corresponding to the delivered purchase opportunity, to maintain information such as “customer order history” for future use in presenting offers and facilitating remote ordering. *See also id.*, [0066]-[0067].

238. In addition, the combination of Kennewick with Hao also renders obvious this limitation. Hao discloses maintaining a “shopper profile/preferences data store” that stores records of user transactions to create a “*past transaction history*” “used to increase accuracy of future transactions.” EX1007, [0033], [0041]; *see also id.*, [0009] (“ambiguities associated with the natural language input can be

automatically resolved based upon previously recorded shopper preference, a shopper profile, *purchase history*, etc.”); [0027] (“*historical shopping information* . . . can be used to determine a shopper preferred brand and size of an otherwise ambiguous item”). For example, Hao discloses that the user may interact with a delivered purchase opportunity, e.g., a presented shopping cart having one or more items available for purchase, by making “*a subsequent request*” to “finalize the e-commerce transaction” and purchase the presented items, which corresponds to “*completing a transaction related to the purchase opportunity.*” *Id.*, [0034], [0040]-[0041]. Because the user interacts with the purchase opportunity through the interface provided by a “client device” (“the electronic device”), the user’s request to “finalize the e-commerce transaction” involves “*the electronic device interact[ing] with the purchase opportunity*” and “*completing a transaction related to the purchase opportunity.*” Hao discloses that the system tracks and stores this interaction pattern to maintain a record of the user’s “past transaction history.” *Id.*, [0021], [0041].

239. A POSITA would have found it obvious to modify Kennewick’s system to track interaction patterns associated with the delivered purchase opportunity, to the extent Kennewick does not already disclose it, by recording “historical shopping information” including user purchases as taught by Hao. EX1007, [0021], [0027], [0040]-[0041]. For instance, Kennewick discloses each domain may “collect, store

and use specific profile information, as required for optimal operations,” which in the context of remote ordering and e-commerce includes the user’s “order history.” EX1006, [0031], [0066]-[0067]. Thus, in the combined system in which Kennewick’s system is modified to implement the functionality of Hao to select and deliver purchase opportunities, a POSITA would have found it obvious to also incorporate, to the extent not explicitly disclosed by Kennewick, Hao’s teaching of tracking an interaction pattern associated with delivered purchase opportunities to maintain a record of goods or services purchased by the user through Kennewick’s remote ordering service that can be “used to increase accuracy of future transactions.” EX1007, [0041]. A POSITA would have found this modification to be obvious as the use of a known technique (e.g., tracking a user’s “historical shopping information”) to improve similar devices in the same way, by enabling Kennewick’s system to use a user’s tracked “historical shopping information” to resolve ambiguities and “increase accuracy of future transactions.” *Id.* The combination of Kennewick and Hao thus renders limitation [1.f] obvious.

[1.g] “building or updating, by the one or more processors, a user-specific profile based on the tracked interaction pattern; and”

240. Kennewick in combination with Hao renders this limitation obvious. As discussed above, Kennewick discloses that the main speech processing unit (e.g., “*one or more processors*”) builds and maintains a “*user-specific profile*” for each user containing user-specific data including the user’s “history of questions asked,

session histories,” and “customer order history.” EX1006, [0016], [0031], [0067]. For example, FIG. 1 reproduced below illustrates that the main speech processing unit 98 maintains “User Profile[s] 110.”

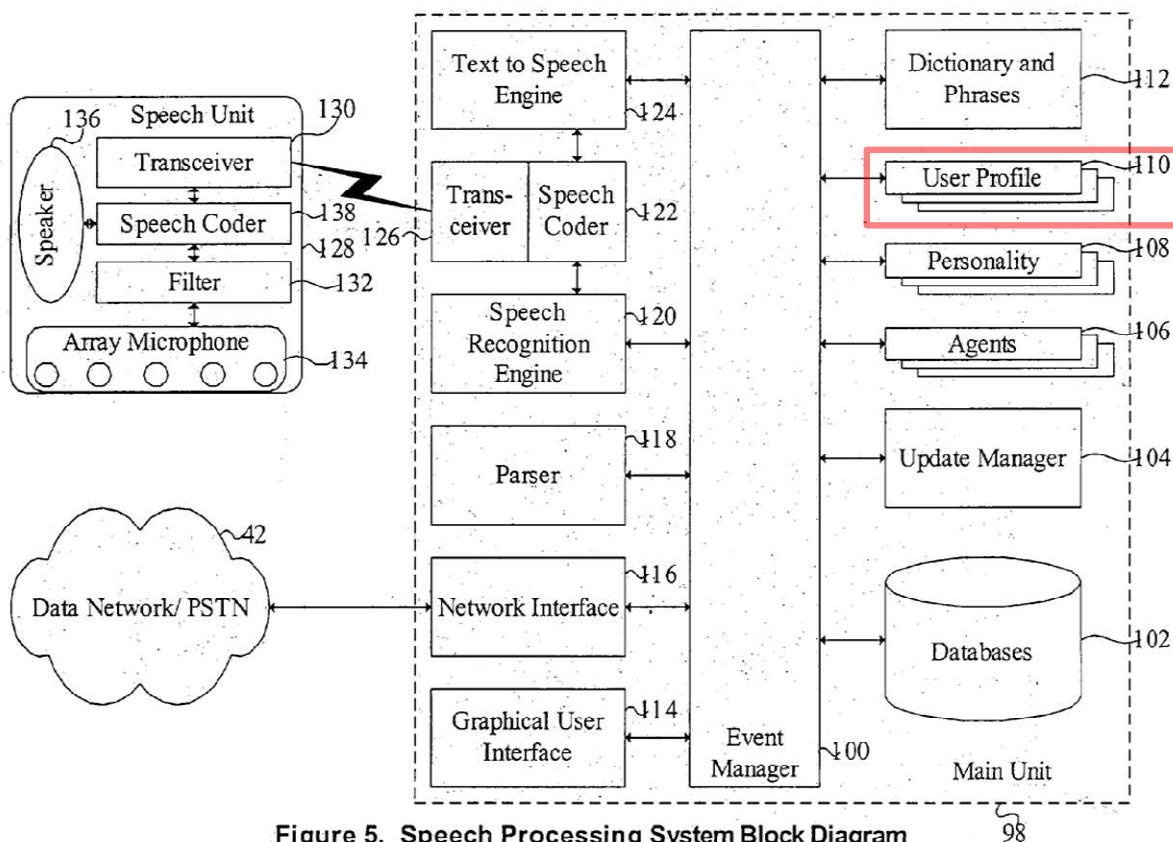


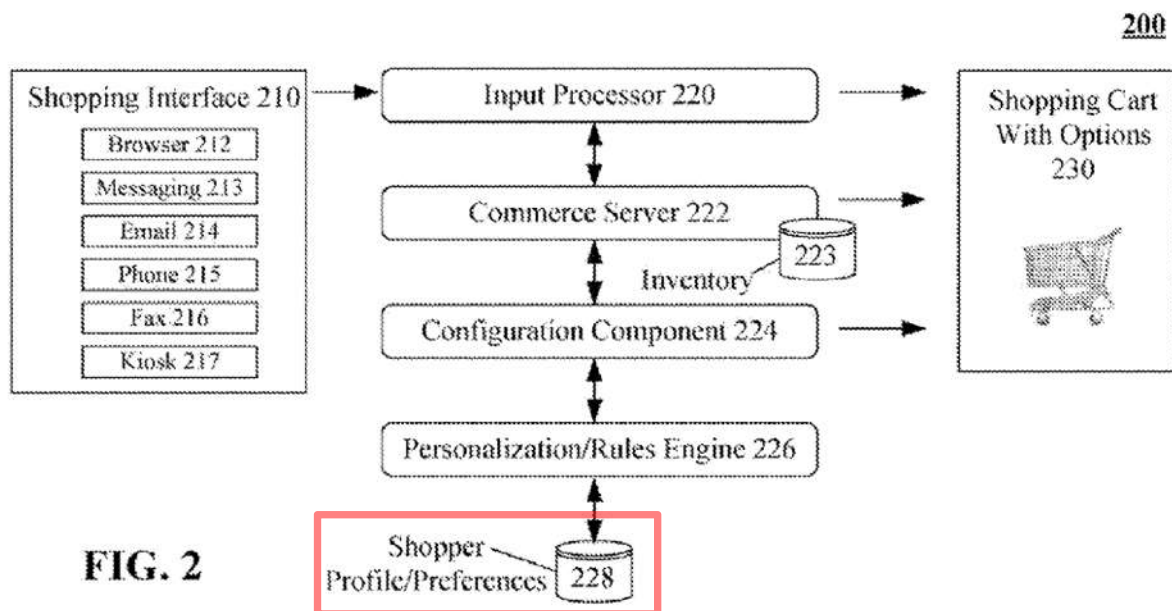
Figure 5. Speech Processing System Block Diagram

241. As discussed above, a POSITA would have understood that a user’s past “order history” corresponds to a “*tracked interaction pattern*” associated with the user, because it describes the user’s prior interactions with previously-presented purchase opportunities that include “a subsequent request” and “complet[ion] [of] a transaction related to the purchase opportunity.” A POSITA would have found it obvious for Kennewick’s “personal profile information for each user”

(corresponding to the “*user-specific profile*”) that includes the user’s “past behavior or history” to include the user’s “order history” information, given Kennewick’s disclosure of using “customer order history” to “facilitate[] remote ordering.” *See id.*, [0031], [0067], [0143]; *see also id.*, [0120] (“User specific data, parameters, session information, location data and history information determining the behavior of agents are stored in one or more user profiles 110.”). In addition, a POSITA would have found it obvious that when the user makes a new purchase when interacting with a presented purchase opportunity, the main speech processing unit “*build[s] or updat[es]*” the user’s user-specific profile based on the tracked interaction pattern of the user’s purchase, so that the user’s “order history” is kept up to date. *See id.*, [0031] (“[i]nformation in the profile may be added and updated automatically as the user uses the system”).

242. In addition, the combination of Kennewick with Hao also renders obvious this limitation. Hao discloses storing a user’s “past transaction history” as part of “shopper specific information maintained within a shopper profile/preferences data store,” and using it to “increase accuracy of future transactions.” EX1007, [0033], [0041]. The system builds and updates the “shopper specific information” using “automatically determined shopper data . . . and historical e-commerce transaction data.” *Id.*, [0033]. Thus, Hao discloses a “*user-specific profile*” (i.e., a “shopper profile” having “shopper specific information”)

built or updated based on “*tracked interaction pattern*[s],” e.g., the user’s “historical e-commerce transaction data” / “past transaction history.” *Id.*, [0033], [0041]. For example, FIG. 2 of Hao (reproduced below) illustrates the system including a “Shopper Profile/Preferences” data store 228 storing the shopper specific information that it accesses through a “Personalization/Rules Engine 226.” *Id.*, FIG. 2.



243. A POSITA would have understood that in the combined system, it would be obvious to modify Kennewick’s system to build and update “user profiles and histories” for each user based on Hao’s disclosure of tracking interaction patterns corresponding to the user’s “historical e-commerce transaction data”/“past transaction history.” A POSITA would have been motivated to make this

combination, to the extent Kennewick does not already disclose it, to make use of a user's "historical e-commerce transaction data" as taught by Hao to select which products or services to present to the user as part of a purchase opportunity. *See* EX1007, [0033], [0041]. By tracking the user's purchasing/order history, the system's ability to select products or services for future purchase opportunities is improved, even when the user's request contains ambiguities. *See also* EX1006, [0032] (describing using "user profile information" to "minimize the interaction with the user required to deliver a response"). 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 (e.g., a user-specific profile containing a record of the user's "order history" to "facilitate[] remote ordering" and "increase accuracy of future transactions"). EX1006, [0016], [0067]; EX1007, [0033], [0041], [0043]. The combination of Kennewick and Hao thus renders limitation [1.g] obvious.

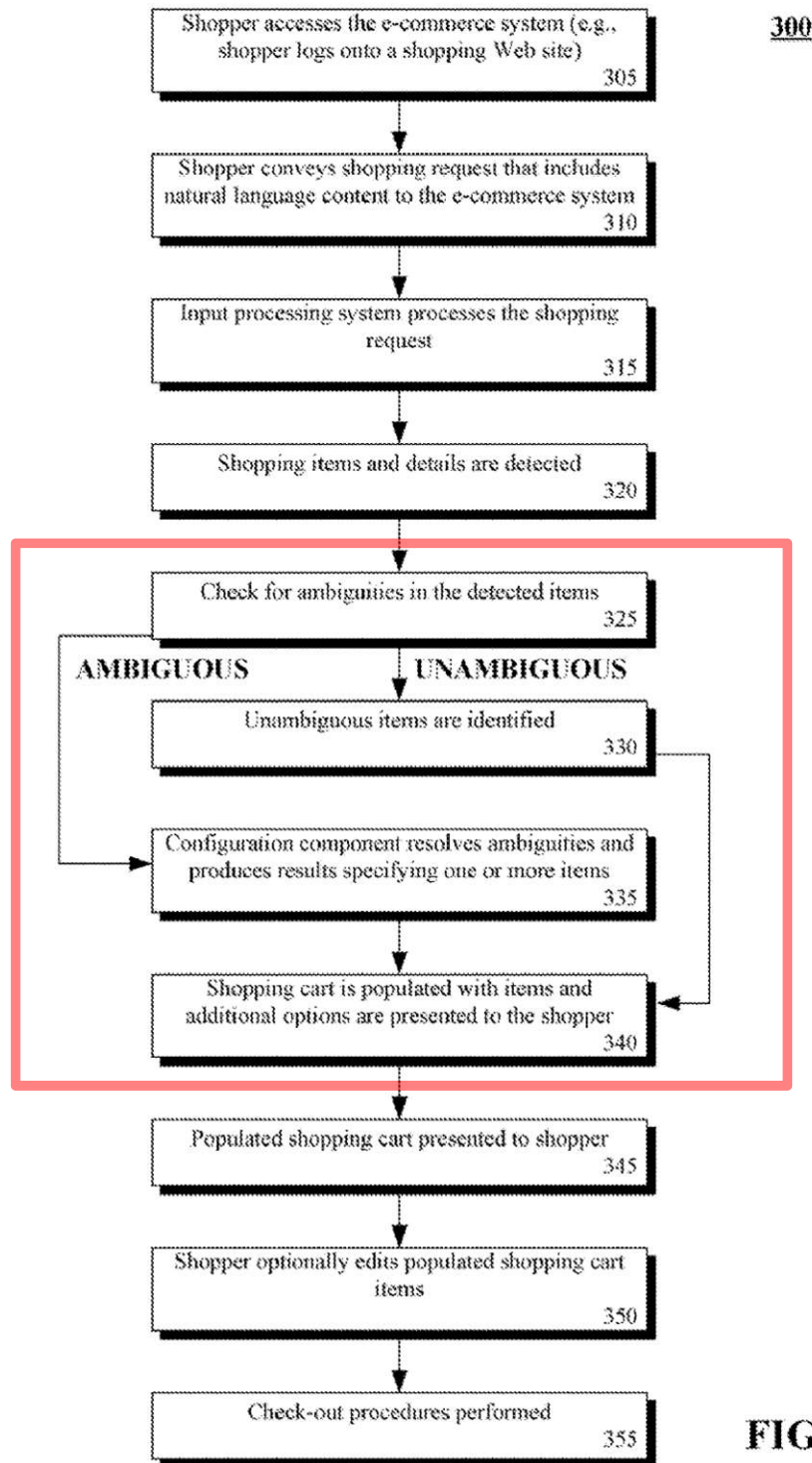
[1.h] "interpreting, by the one or more processors, a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance."

244. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the speech main processing unit (e.g., "***one or more processors***") includes a parser 118 that performs "parsing and interpreting" of each

of the user's utterances, to transform the recognized words and phrases from each utterance "into complete commands and questions using data supplied by the agents." EX1006, [0028], [0123]. As part of this parsing and *interpreting*, Kennewick's parser uses "user profile information" to "minimize the interaction with the user required to deliver a response." *Id.*, [0032], [0143] ("Commands and questions are interpreted, queries formulated, responses created and results presented can be based on the user's personal or user profile 110 values."). For example, Kennewick discloses applying user-specific profile information ("user profile" and "history") to respond to user queries for "offers and promotions for goods and services," and using "customer preferences [and/or] customer order history" to "facilitate[] remote ordering" of goods and services. *Id.*, [0066]-[0067]. A POSITA thus would have understood that Kennewick discloses "*interpreting*" a user's natural language utterance "*using the user-specific profile*" associated with the user, e.g., by using profile data built and/or updated based on the user's order history and past transactions to interpret the utterance. In addition, a POSITA would have found it obvious that Kennewick selects a purchase opportunity based on the natural language utterance, by providing offers and promotions for goods and services responsive to the natural language utterance. *See* limitation [1.d]. A POSITA would also have understood that the natural language utterance and purchase opportunity are a "*subsequent natural language utterance*" and a

“subsequent purchase opportunity,” as they relate to an interaction that is processed by the system using user profile information built based on the user’s past interactions with the system involving a previous natural language utterance and a previous purchase opportunity.

245. To the extent that Kennewick does not explicitly disclose “interpreting . . . a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance,” the combination with Hao renders it obvious. As discussed above, Hao discloses using a user’s “past transaction history” stored as part of a “shopper profile” to “increase accuracy of future transactions.” EX1007, [0033], [0041]. For example, Hao’s “input processor” includes a natural language understanding (NLU) engine that, when interpreting a user’s request, performs a “check for ambiguities” as shown in element 325 of FIG. 3 (reproduced below).

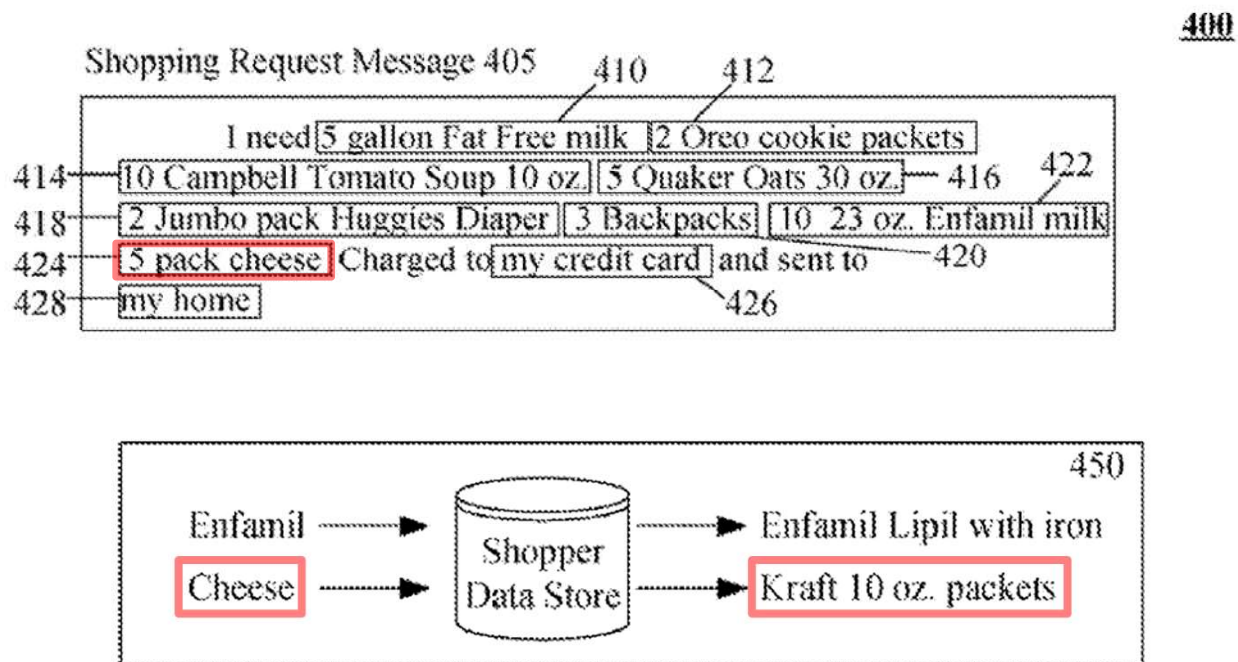


Id., FIG. 3, [0039]. If the input processor identifies an ambiguity, a “configuration component” attempts to “automatically resolve[]” the ambiguity using information

from the user's shopper profile (i.e., "***user-specific profile***"), including the user's "purchase history." *Id.*, [0009], [0039]. Upon resolving the ambiguity, the configuration component "produces results specifying one or more items" corresponding to the user's request and presents the items to the shopper as part of a "shopping cart." *Id.*, FIG. 3 (element 335). The "produce[d] results specifying one or more items" are a "***purchase opportunity***" selected based on the natural language utterance "***interpret[ed] . . . using the user-specific profile,***" because the system selects the results based on the received natural language utterance having an ambiguity that is resolved using the past transaction information from the user's profile. In addition, because the system interprets the user's natural language request using "past transaction history" associated with previous requests, a POSITA would have understood that the user's request is a "***subsequent natural language utterance***" and that the "produce[d] results specifying one or more items" "presented to the shopper" are a "***subsequent purchase opportunity.***" *Id.*, [0039], [0041] ("The method 300 can be repeated for other e-commerce transactions as desired."), FIG. 3. Thus, Hao discloses "***interpreting a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance.***"

246. Hao further provides an example of how the system may interpret an ambiguous user request corresponding to a "***subsequent natural language***

utterance” using a “*user-specific profile*” (“shopper profile”) to select a purchase opportunity in relation to FIG. 4, portions of which are reproduced below.



EX1007, FIG. 4 (reproduced portion). As previously discussed, the user request may correspond to a “*natural language utterance*.” *See id.*, claims 8 (user input is “speech input in a natural language format”), 20 (user interacts with system through “voice only interface”). In addition, the user request may be a “*subsequent*” natural language utterance because the system may interpret it using a user-specific profile (“shopper profile”) built and/or updated based on user interaction patterns associated with previous purchase opportunities presented in response to the user’s previous natural language utterances. *See id.*

247. As shown in the example of FIG. 4, the user's request includes a request for "5 packs of cheese." *Id.*, [0042]. Hao's input processing component identifies and categorizes the recognized words or phrases from the user's request to determine that certain portions of the request, such as "5 packs of cheese," correspond to "ambiguous items" where "[b]rands and/or other purchase specifics" are "initially unknown." *Id.*, [0043]-[0045]. To resolve these ambiguities, Hao's input processing component uses information from the user's user-specific profile, including "purchase history, preference, and profile information of the shopper" to interpret the request for "cheese" as referring to "KRAFT 10 oz. packets," which it then places in the shopping cart presented to the user. *Id.*, [0042], [0046], FIG. 4.

248. Thus, Hao discloses interpreting a subsequent natural language utterance using the user-specific profile to select a subsequent purchase opportunity based on the subsequent natural language utterance, such as by interpreting the user's natural language input (which may be received verbally as an utterance) containing a request for "5 pack cheese" as referring to "KRAFT 10 oz. packets," and populating a shopping cart with the identified item for presentation to the user as a purchase opportunity. *See also* EX1007, [0024] (e-commerce server may be "an Interactive Voice Response (IVR) system"), [0037] ("The shopping request can be provided using online forms, e-mail, fax, voice over a voice communication channel . . . and the like"), claim 20 (shopping interface "is a voice only interface

for interacting with an automated voice response system over a real-time communication channel”).

249. While Hao describes FIG. 4 as illustrating an example of “free form text processing during an e-commerce transaction,” a POSITA would have understood that the example shown in Hao’s FIG. 4 applies equally to natural language inputs received as speech utterances. *See* EX1007, [0042]. For example, Hao discloses that Hao’s e-commerce shopping interface enables shoppers to “submit requests to a commerce server in a natural language” via different modalities, such as “verbal and/or text based” requests. *Id.*, [0009], [0010]. However, regardless of the input modality in which natural language input is received, the input processing component converts any non-text input into text for processing by the “natural language understanding (NLU) engine 132.” *Id.*, [0010], [0022]. For example, the input processing component includes an “automatic speech recognition (ASR) engine 134” that processes speech input to “convert the shopper provided utterances into *text*, which is further processed by the natural language understanding engine,” while “an optical character recognition (OCR) engine 136” pre-processes image-based input to extract text content that is then processed by the NLU engine. *Id.*, [0022], claim 15. Thus, a POSITA would have understood that the “free form text processing” described in relation to Hao’s FIG. 4 applies equally to user’s free form shopping inputs received directly as text as well as those received

via speech that are converted to text through ASR prior to being processed by the NLU engine.

250. A POSITA would have found it obvious to combine Kennewick and Hao by modifying Kennewick’s marketing and remote ordering services to incorporate Hao’s teaching of interpreting a subsequent natural language utterance using a user-specific profile (e.g., “past transaction history” from a “shopper profile”) to select a subsequent purchase opportunity. For example, Kennewick explains that to more accurately interpret and respond to user requests, the system may make “significant use of context, prior information, domain knowledge, and user specific profile data.” EX1006, Abstract, [0007], [0011], [0118]. Indeed, Kennewick recognizes that “some questions” may be ambiguous, and “can only be interpreted in the context of . . . the user’s history of interests and preferences.” *Id.*, [0007]. This may include use of customer-specific information such as “customer preferences” and “customer order history” to facilitate “[r]emote ordering and payment for goods and services.” *Id.*, [0067]. Hao, in turn, also recognizes that a shopper’s “past transaction history can be . . . used to increase accuracy of future transactions,” and describes specific techniques for interpreting and resolving ambiguities in a shopper’s current request “by using shopper specific information maintained within a shopper profile/preferences data store,” such as “items that have previously been purchased by [the] shopper.” EX1007, [0033], [0040]-[0041]. By

resolving these ambiguities, Hao's system can place the appropriate items in the user's electronic shopping cart for the user to purchase, without needing to solicit additional information from the user. *See id.*, [0039]-[0041], [0046].

251. A POSITA in possession of Kennewick and Hao would therefore have been motivated to incorporate Hao's techniques for interpreting and resolving ambiguous user requests based on user-specific information such as "past transaction history" to select and present purchasable items to the user. The modification would enhance and improve how Kennewick uses "user specific profile data" such as "customer preferences" and "customer order history" to facilitate remote ordering of goods or services, by enabling Kennewick's system to interpret user requests and present a purchase opportunity even when the request is "incomplete, ambiguous or subjective." *See* EX1007, [0041]; EX1006, [0002], [0007], [0067], [0018]. By resolving these ambiguities based on the teaching of Hao, the combined system can identify "one or more items" to present to the user as part of selected purchase opportunity, without having to solicit the user for additional information, thus "minimiz[ing] the interaction with the user required to deliver a response." EX1007, [0039], [0045]-[0047]; EX1006, [0032]. This modification is a combination of prior art elements disclosed in Kennewick and Hao (e.g., Kennewick's general disclosure of using "customer order history" to facilitate remote ordering, and Hao's more specific disclosure of using "past transaction history" to resolve ambiguities and

select items for purchase opportunities) according to known methods to yield predictable results. The combination of Kennewick and Hao thus renders limitation [1.h] obvious.

3. Claim 2

[2.pre] “The method of claim 1, wherein the purchase opportunity relates to an item, the method further comprising:”

252. Kennewick in combination with Hao renders the preamble obvious. As discussed above, Kennewick in combination with Hao renders obvious the method of claim 1. In addition, Kennewick and Hao each disclose that “the purchase opportunity” may relate to “*an item*.” For example, Kennewick discloses presenting purchase opportunities, e.g., “interactive offers,” allowing for “[r]emote ordering and payment for goods” corresponding to “shopping items” or “items to purchase.” EX1006, [0066]-[0067], [0070]. Similarly, Hao also discloses that “*the purchase opportunity relates to an item*” because it discloses “extract[ing] semantics from the [user’s] input that relates to at least one *item* available for sale,” and presenting a purchase opportunity to the user (e.g., by placing the item in the user’s shopping cart) if “an unambiguous match between extracted semantics and inventory items occurs.” EX1007, [0032]-[0033]. The combination of Kennewick and Hao therefore renders limitation [2.pre] obvious.

[2.a] “receiving, by the one or more processors, a second natural language utterance;”

253. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the main speech processing unit (e.g., “***one or more processors***”) receives natural language utterances. *See, e.g.*, EX1006, [0121]-[0123] (describing “receiv[ing] the utterances of the user” as “coded speech”); [0155] (describing performing speech recognition “for each utterance”). Because Kennewick discloses receiving and processing multiple natural language utterances, a POSITA would have understood that the received natural language utterances may include a “***second natural language utterance.***” *See also id.*, [0007], [0177] (describing interpreting “subsequent commands” based on “previous commands”). The combination of Kennewick and Hao therefore renders limitation [2.a] obvious.

[2.b] “determining, by the one or more processors, that the second natural language utterance comprises a request to purchase the item; and”

254. Kennewick in combination with Hao renders this limitation obvious. As discussed above, Kennewick’s system enables users to perform “[r]emote ordering and payment for goods,” which as discussed above, may relate to an “item.” EX1006, [0067]. In addition, Hao discloses that when items are presented to the user, the user can select from “shopping interactive options” including an option to “finalize . . . an e-commerce transaction involving a purchase of the items presented in the shopping cart.” EX1007, [0011], [0041]. A POSITA would have understood

that in the combined system, the user may select an option to finalize an e-commerce transaction to purchase a presented item as taught by Hao, corresponding to “*a request to purchase the item*” associated with the purchase opportunity, to perform “[r]emote ordering and payment for goods” as taught by Kennewick.

255. Kennewick and Hao both disclose that the user may interact with the system through an interactive speech interface that processes the user’s natural language queries or commands. *See* EX1006, [0101] (“Virtually any service, involving the transfer of data or transmission of speech and video, can be supported through the natural language speech interface”); EX1007, claim 20 (e-commerce interface may be “a *voice only interface*”). A POSITA thus would have found it obvious for the user’s request to finalize the e-commerce transaction and purchase the item to be in the form of a natural language utterance, e.g., a “*second natural language utterance*,” that is received by the system through the speech or “voice only” interface.

256. A POSITA would have found it obvious to modify Kennewick, to the extent it does not explicitly disclose it, to enable the user to select an option to finalize an e-commerce transaction to purchase a presented item as taught by Hao, as it provides a straightforward way for Kennewick’s system to enable remote ordering of items. For example, Kennewick’s parser interprets received utterances and “transforms them into complete commands and questions,” and a POSITA

would have found it obvious for the parser in the combined system to interpret the second natural language utterance as a command to finalize the transaction to purchase the item as taught by Hao, to complete a remote ordering transaction. *See* EX1006, [0067], [0123]. The parser would thus “**determin[e] . . . that the second natural language utterance comprises a request to purchase the item.**” A POSITA would have found such a modification to be obvious, as it applies a known technique (e.g., the user selecting an option to purchase an item associated with a presented purchase opportunity through speech via a voice only interface) to a known device (e.g., Kennewick’s speech interface that enables remote ordering of products or services) ready for improvement to yield predictable results (e.g., the user performing remote ordering by selecting an option to purchase an item associated with a presented purchase opportunity via a speech utterance). The combination of Kennewick and Hao therefore renders limitation [2.b] obvious.

[2.c] “completing, by the one or more processors, a purchase transaction of the item based on the second natural language utterance.”

257. Kennewick in combination with Hao renders this limitation obvious. For example, Kennewick’s main speech processing unit “facilitates remote ordering” enabling the user to complete a purchase transaction (“ordering and payment”) for an item through a “natural language speech interface.” *See* EX1006, [0067], [0101]. Similarly, Hao discloses that the user may “finalize the e-commerce transaction” to purchase an item through a “voice only interface.” *See* EX1007,

[0041], claim 20. The system completes the purchase transaction, which may including sending “payment . . . for purchased items” and shipping the items “to an address designated by” the user. *Id.*, [0022]. Thus, Kennewick and Hao render obvious “***completing . . . a purchase transaction of the item***” based on received user voice inputs (e.g., a “***second natural language utterance***”). The combination of Kennewick and Hao therefore renders limitation [2.c] obvious.

4. Claim 4: “The method of claim 1, wherein delivering the selected purchase opportunity via the electronic device comprises: audibly presenting, by the electronic device, a natural language conversational purchase opportunity.”

258. Kennewick in combination with Hao renders obvious claim 4. First, as discussed for limitation [1.e], Kennewick and Hao disclose delivering the selected purchase opportunity via the electronic device. A POSITA would have found it obvious for the system to “***audibly present[]***” the purchase opportunity as a “***natural language conversational purchase opportunity***,” given Kennewick’s disclosure presenting responses to the user 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). EX1006, [0024], [0121], [0124]. For example, Kennewick discloses audibly presenting responses delivered to the user through a speaker. *Id.*, [0121], [0124]. Kennewick further describes techniques for “mak[ing] the responses to user’s questions and commands seem more natural” and conversational. *See, e.g., id.*, [0035], [0179]-[0180]. A POSITA would have found

it obvious for Kennewick's system to present purchase opportunities in this way, i.e., audibly and in a natural language conversational manner.

259. Similarly, Hao discloses that after the system determines “one or more shopping items . . . from the shopping request,” it presents a “shopper specific shopping cart . . . populated with items ascertained from the shopping request” to the user. EX1007, [0038], [0040]. A POSITA would have found it obvious for the system to present the shopping cart “*audibly*” as “*a natural language conversational purchase opportunity*,” based on Hao's disclosure that the e-commerce interface may be a “voice only interface.” EX1007, claim 20. In the combination discussed above in which Kennewick's system is modified to select a purchase opportunity based on the teachings of Hao, it would have been obvious to also present the selected purchase opportunity “*audibly*” as “*a natural language conversational purchase opportunity*” as taught by Kennewick and Hao, which aligns with Kennewick's stated objective of creating a “telematics applications” that can “accommodate commands and queries from a wide range of domains.” See EX1006, Abstract, [0003]-[0004]. The combination of Kennewick and Hao therefore renders claim 4 obvious.

5. Claim 5

[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:”

260. Kennewick in combination with Hao renders the preamble obvious. As discussed above, Kennewick discloses that the one or more physical processors of the “main speech processing unit” are programmed with computer program instructions, which when executed cause the one or more physical processors to perform the method. For example, Kennewick’s main speech processing unit contains various software modules, e.g., “a speech recognition module,” “a parser,” and “a text to speech engine,” as well as “agents” that are “packages of executable code.” EX1006, [0013], [0017]. These software modules and packages of executable code correspond to the claimed “**computer program instructions**” that, when executed by the one or more processors of the main speech processing unit, cause the one or more processors to perform the method steps.

261. In addition, Kennewick discloses that the software modules (i.e., “**computer program instructions**”) implemented on the main speech processing unit includes a “parser 118” and “agents 106” which operate together to “**interpret a natural language utterance.**” See EX1006, FIG. 5. For example, the parser 118 interprets the recognized words and phrases of the natural language by “transform[ing] them into complete commands and questions using data supplied by

the agents.” *Id.*, [0123], [0155]; *see also id.*, [0163] (describing that the “parser and domain agent use” domain-specific information “to determine the most plausible interpretation” of an utterance). The parser and agents of the main speech processing unit thus correspond to the claimed “*conversational language processor configured to interpret a natural language utterance.*”

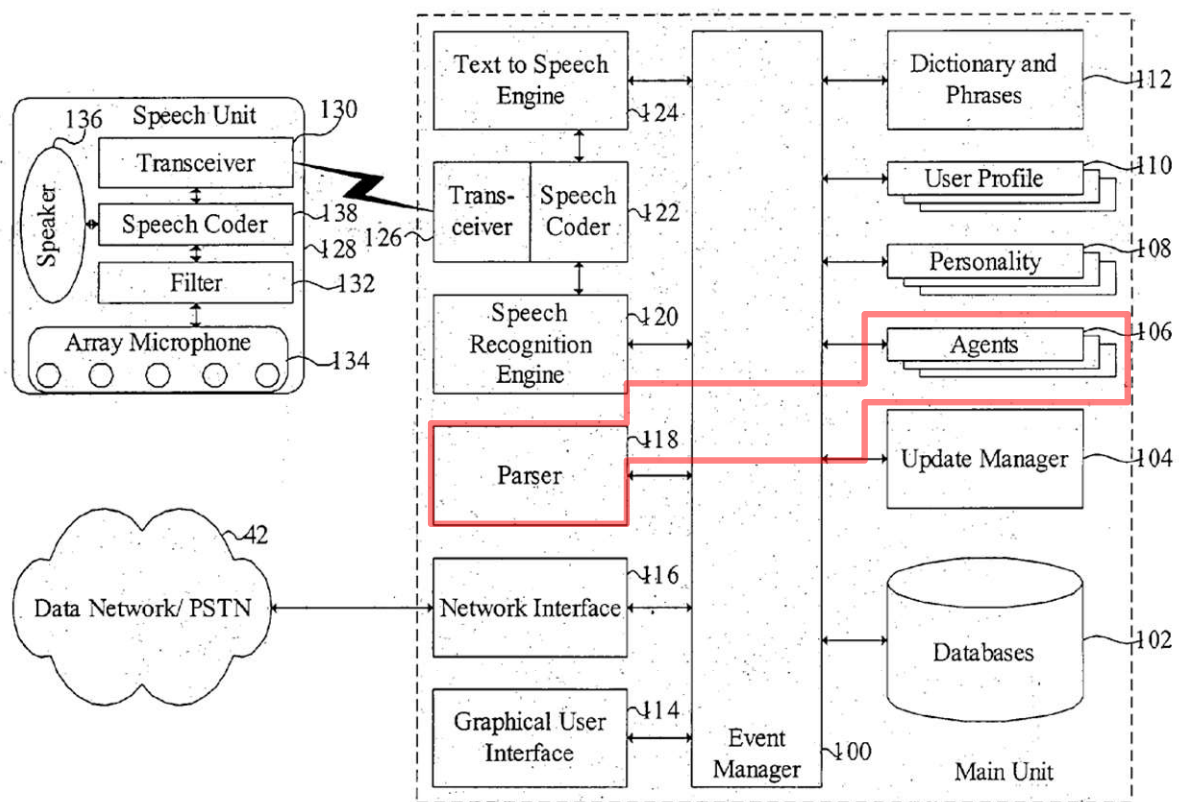


Figure 5. Speech Processing System Block Diagram

262. This mapping is consistent with the disclosure of the challenged patent specification, which describes that the “conversational language processor” includes a “context determination module 130” and “one or more agents 135.” EX1001, 4:31-37, FIG. 1. Kennewick’s “parser 118” is analogous to the ’758 patent’s context

determination module, being configured to “determine a context for an utterance” as discussed above. EX1006, [0160]. Also, Kennewick’s agents 106, which provide “packages or modules of functionality” to “receive, process and respond to user questions, queries and commands” for “a specific domain,” are analogous to the ’758 patent’s “one or more agents 135,” which each provide a “package of functionality for a specific application that receives, processes, and responds to user questions, queries, and/or commands.” EX1006, [0017]; EX1001, 5:9-14 (“agents 135 may model various domains”), 14:35-40. Thus, a POSITA would have recognized Kennewick’s parser 118 and agents 106 as corresponding to the claimed *“conversational language processor.”* The combination of Kennewick and Hao thus renders limitation [5.pre] obvious.

[5.a] “providing, by the one or more processors, the words or phrases as an input to the conversational language processor; and”

263. Kennewick in combination with Hao renders this limitation obvious. As I discussed above, Kennewick discloses that the main speech processing unit (e.g., the *“one or more processors”*) includes a “speech recognition module,” “a parser,” and agents, where the parser and agents correspond to the claimed *“conversational language processor.”* EX1006, [0013], [0017]. As discussed above for limitation [1.b], the main speech processing unit receives words or phrases recognized from the natural language utterance as an output of the speech recognition engine, because the speech recognition engine sends the words and

phrases to event manager 100, a component of the main speech processing unit. *See also id.*, [0119], [0123], [0155]. The main speech processing unit then “pass[es]” these words or phrases “to the parser 118 for interpretation” via the “event manager 100.” *Id.*, [0119], [0155]. As these modules are all part of the main speech processing unit, a POSITA would have found it obvious that the main speech processing unit “*provides the words or phrases as input*” to the parser 118, which, as discussed above, is part of the claimed “conversational language processor.” The combination of Kennewick and Hao thus renders limitation [5.a] obvious.

[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.”

264. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the main speech processing unit (e.g., the “*one or more processors*”) performs, for “each user utterance including a question or query or set of questions or queries, . . . parsing and interpreting the question or query.” EX1006, [0028]. As discussed above, the parser 118 and agents 106, which form the claimed “*conversational language processor*,” perform this parsing and interpretation. For example, Kennewick discloses that the parser 118 receives the recognized words or phrases of the natural language utterance and interprets them by “transform[ing] them into complete commands and questions using data supplied by the agents.” *Id.*, [0123], [0155]; *see also id.*, [0162] (the parser “formulate[s] a question or command

in the regular grammar used by agents”), [0163] (describing that the “parser and domain agent use” domain-specific information “to determine the most plausible interpretation” of an utterance). The formulated “command or question” meets the claimed “*interpretation of the natural language utterance, based on the words or phrases*,” and is used by an invoked domain agent to query “one or more local and/or network data sources” to obtain one or more results to be presented to the user as a response. *See id.*, [0028], [0165]. As the parser and agents are modules within the main speech processing unit, a POSITA would have found it obvious that the main speech processing unit “*receives*,” e.g., via the event manager 100, the interpretation as “*an output of the conversational language processor*,” e.g., as an output of the parser 118 which as discussed above is part of the claimed “conversational language processor.” *See also id.*, [0119] (“The event manager 100 may mediate interactions between other components of the invention”), [0127].

265. In addition, Kennewick renders obvious that the “*the purchase opportunity is selected based further on the interpretation of the natural language utterance*.” For example, Kennewick discloses that the parser’s processing of an utterance includes “evok[ing] the correct agent to process the question or command” (“*the interpretation of the natural language utterance*”), where the agent formulates “one or more queries . . . sent to local and/or network information sources” based on the “question [or] command” (the claimed “interpretation”), and

formats a response to the query “for presentation” to the user. EX1006, [0164]-[0165], [0167]-[0168], [0178]. A POSITA would have found it obvious that in contexts where the natural language utterance relates to an e-commerce context, such as a “query . . . for offers and promotions for goods and services,” the main speech processing unit evokes the appropriate domain agent which responds to the user with the “purchase opportunity” (e.g., “offers and promotions”) selected based on the results of the queries formulated by the agent based on the interpretation of the utterance. *See id.*, [0066].

266. To the extent that Kennewick does not explicitly disclose that “*the purchase opportunity is selected based further on the interpretation of the natural language utterance*,” the combination with Hao renders it obvious. Hao’s input processing system includes a “natural language understanding (NLU) engine 132” that extracts “semantic meanings” from the recognized text of the natural language utterances, which meets the claimed “*the interpretation of the natural language utterance*.” EX1007, [0022], [0028], [0043]. Hao’s input processing system compares the extracted semantic meanings “to inventory items available to the shopper 110 for purchase from at least one merchant system,” and adds matching items “to an electronic shopping cart” to form the claimed “purchase opportunity.” *Id.*, [0022], [0043]. Thus, Hao discloses that “*the purchase opportunity is selected based further on the interpretation of the natural language utterance*,” because it

selects the purchase opportunity by comparing the “semantic meanings” of the interpretation with available inventory items.

267. In the combined system, a POSITA would have found it obvious to modify Kennewick’s system, to the extent Kennewick does not explicitly disclose how the purchase opportunity is selected, to select the purchase opportunity based on the interpretation of the natural language utterance, by comparing the interpretation with available inventory items to identify items to be included in a purchase opportunity as taught by Hao. A POSITA would have been motivated to make this combination to implement a straightforward way for Kennewick’s system to select products or services to present to the user responsive to user requests for offers or promotions. For example, 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 interactive speech interface which presents offers and promotions for goods or services in response to user queries) in the same way, so that Kennewick’s system compares the interpretation of the natural language utterance to available inventory items to select items for a purchase opportunity. 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. The combination of Kennewick and Hao therefore renders limitation [5.b] obvious.

6. Claim 6: “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: (i) interpreting requests related to its domain; and (ii) determining a response to the requests related to its domain.”

268. Kennewick in combination with Hao renders claim 6 obvious. As discussed above, Kennewick’s “parser 118” and “agents 106” meet the claimed “*conversational language processor*” “configured to interpret a natural language utterance.” Kennewick further discloses that the agents 106 of the claimed “*conversational language processor*” include one or more domain agents, each of which includes “the behavior and data required for a specific area of functionality.” EX1006, [0018]; *see also id.*, [0160] (“The context of a question or command may determine the domain and thereby, the domain agent 156, if any, to be invoked”). For example, FIG. 6, reproduced below, shows that the agents 106 include multiple domain agents 156.

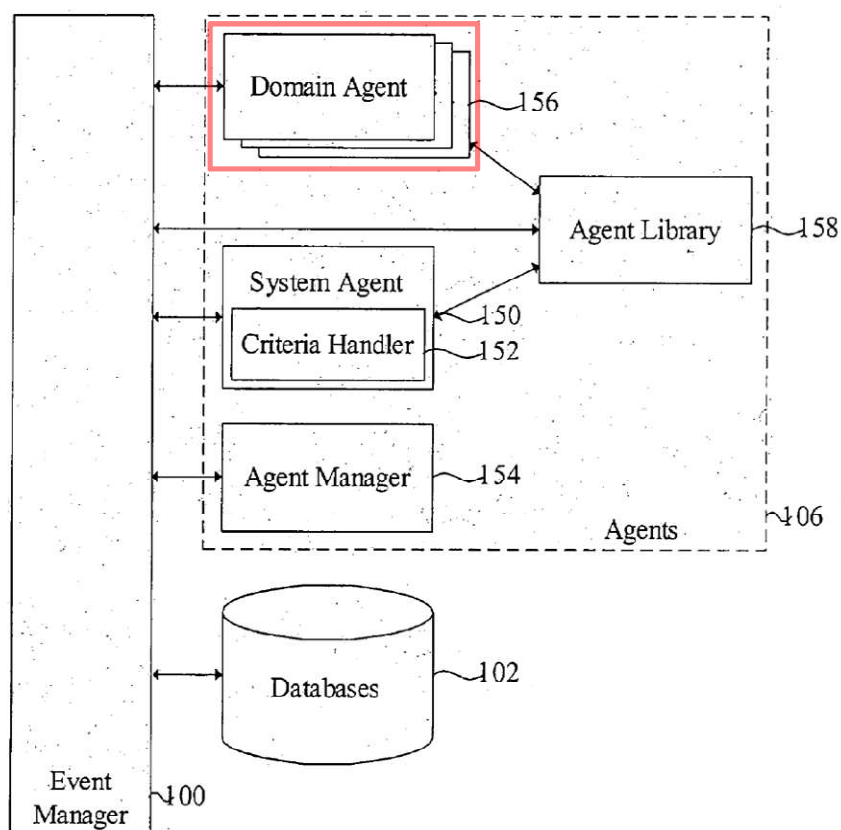


Figure 6. Agent Architecture

269. Kennewick discloses that each domain agent includes “the behavior and data required for a specific area of functionality” for assisting the parser in interpreting requests relating to its domain. EX1006, [0018]. For example, Kennewick discloses that the “parser and domain agent [may] use” domain-specific information “to determine the most plausible interpretation” of an utterance and resolve potential ambiguities. *Id.*, [0163] (describing that the parser “can use services of the domain agent 156 to look up tables in the databases 102” to interpret an utterance); *see also id.*, [0029] (interpreting an utterance “may be performed using

. . . domain specific information”), [0123] (the parser generates the interpretation “using data supplied by the agents”). Thus, Kennewick discloses that the domain agent is configured to “*assist in . . . interpreting requests related to its domain*” because it provides data to the parser for interpreting user requests related to its domain. *See also id.*, [0163] (describing a travel services domain agent using “flight information in the database and network information” to assist the parser in interpreting an utterance).

270. In addition, each domain agent further “*assists in . . . determining a response to the request relating to its domain*” by, when invoked, “create[ing] one or more queries to one or more local or external information sources” related to its domain, and “create[ing] a satisfactory response to [the] question” based on the query results. EX1006, [0167], [0178]. The domain agent creating a response based on querying information sources related to its domain meets the claimed “*determining a response to the request relating to its domain.*”

271. As discussed above for claim 5, a POSITA would have found it obvious in the combined system to modify Kennewick to select the purchase opportunity based on the interpretation of the natural language utterance as taught by Hao. The selected purchase opportunity meets the determined “*response to the request.*” In this combination, it would have been obvious for an invoked domain agent relating to an e-commerce service, such as a “fast food ordering agent” or “remote ordering”

agent, to select the purchase opportunity as taught by Hao, to “*determin[e] a response to the [user’s] request related to [the invoked] domain.*” See EX1007, [0022]; EX1006, [0018], [0066]-[0067]. A POSITA would have found it obvious to make this modification, as Kennewick discloses that 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 and promotions” for goods or services as a response to the user’s requests, using the teachings of Hao to select the goods or services, as discussed for claim 5. See EX1006, [0017], [0066]. The combination of Kennewick and Hao therefore renders claim 6 obvious.

7. Claim 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.”

272. Kennewick in combination with Hao renders claim 7 obvious. As discussed above, Kennewick discloses “*one or more domain agents,*” where each domain agent “include[s] the behavior and data required for a specific area of functionality” or “specific domain[.]” EX1006, [0018], [0127]. These include domains for travel services, entertainment scheduling, and remote ordering and payment for goods such as fast food, which are examples of “*electronic commerce domain[s].*” *Id.*, [0018], [0067]. Thus, Kennewick discloses that the domains in which the one or more domain agents are configured to assist comprises an electronic commerce domain.

273. In addition, the combination of Kennewick with Hao also renders this claim obvious. For example, Hao discloses an electronic commerce or “e-commerce system” which processes user inputs to return one or more items that the user can then purchase. EX1007, [0010], [0036]. In the proposed combination of Kennewick and Hao discussed above, it would be obvious to combine Kennewick with Hao in which the functionality of Hao’s e-commerce system is implemented as one of the domains in Kennewick’s system. For example, Kennewick discloses maintaining domain agents corresponding to many different areas of functionality. *See, e.g.*, EX1006, [0018], [0060]-[0069], [0130]. Kennewick further discloses that the domain agents are “executables” that provide “re-distributable packages or modules of functionality,” where the system can add new domain agents “to extend the functionality of the invention to a new domain.” *Id.*, [0017], [0126], [0130]. It would be obvious to implement Hao’s e-commerce functionality, which “may be embedded in a computer program product,” as an agent in Kennewick’s system that receives, processes, and responds to user requests relating to the context of an “*electronic commerce domain*.” *See* EX1007, [0055], claim 10 (Hao’s method is preformed “automatically by at least one machine in accordance with at least one computer program having a plurality of code sections that are executable by the at least one machine”). For example, a POSITA would have found it obvious to implement Hao’s e-commerce functionality as part of Kennewick’s services for

remote ordering of goods and services, e.g., to provide a more robust service that is able to interpret ambiguous user requests using user-specific profile information. A POSITA would have been motivated to make this combination as it provides a straightforward way to “extend the functionality of [Kennewick’s] invention” to better support “remote ordering,” through implementing a domain agent that selects a purchase opportunity to be presented to the user based on the user’s request as taught by Hao, even if the user’s request contains ambiguities. A POSITA would have had a reasonable expectation of success because the combination is a simple combination of prior art elements (Kennewick’s teaching of establishing a domain agent for performing “a specific area of functionality,” and Hao’s teaching of a specific area of functionality in an e-commerce context) according to known methods to yield predictable results (e.g., Kennewick’s system establishing a domain agent for performing e-commerce transactions as taught by Hao). The combination of Kennewick and Hao therefore renders claim 7 obvious.

8. Claim 9: “The method of claim 1, the method further comprising: 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”

274. Kennewick in combination with Hao renders claim 9 obvious. First, Kennewick discloses that the main speech processing unit (e.g., the “*one or more processors*”) uses “*an environmental model to determine environmental*

information.” For example, Kennewick discloses that its system can be deployed on “multiple devices . . . placed at different locations,” where the user can interact with the system through any of these devices. EX1006, [0016]. Furthermore, the system may use “the location of the particular device” through which the user is interacting with the system “as part of the context for the questions asked.” *Id.*, [0016]; *see also id.*, [0065] (describing personalizing providing services based on “location”). Thus, a POSITA would have understood that Kennewick’s system maintains data of the locations of each of the multiple devices, which meets the claimed “**environmental model**,” and uses it to determine the “location of the particular device” used by the user, which meets the claimed “**environmental information**.” *See id.* This is consistent with the disclosure of the challenged patent, which explains that “environmental information” may correspond to a “user location.” EX1001, claim 10. Kennewick further discloses that “**the context for the natural language utterance is determined based further on the environmental information**,” as it discloses that the location of the device is used “as part of the context” for the natural language utterance. EX1006, [0016], [0065].

275. Alternatively, Kennewick also discloses that the main speech processing unit (e.g., the “**one or more processors**”) maintains “**an environmental model to determine environmental information**” in the form of “user profiles and histories” from which “environmental information” corresponding to past user

activity and actions can be determined. See EX1006, [0016], [0031]; see also EX1001, claim 10 (claimed “environmental information” may correspond to “a user activity, or a user action”). Kennewick further discloses that “*the context for the natural language utterance is determined based further on the environmental information*” because it discloses determining the context for the natural language utterance based on the information of the user’s past activity and actions as indicated in the “user profile” and “dialog history.” EX1006, [0160]; see also *id.*, [0143] (“Personal profiles may include information specific to the individual,” including “the history of their interactions with the system.”). The combination of Kennewick and Hao therefore renders claim 9 obvious.

9. Claim 10: “The method of claim 1, wherein the environmental information comprises a user location, a user activity, or a user action.”

276. Kennewick in combination with Hao renders claim 10 obvious. As discussed above, Kennewick discloses an “environmental model” used to determine “*environmental information*.” In addition, Kennewick discloses that the claimed “*environmental information comprises a user location, a user activity, or a user action*.” For example, as discussed above, Kennewick discloses that the environmental information may comprise “the location of the particular device addressed by the user,” which is “*a user location*.” EX1006, [0016]. Alternatively, as discussed above, Kennewick discloses that the environmental information may

comprise a “history of [the user’s] interactions with the system” as determined from a user profile and dialog history, which is “*a user activity, or a user action.*” EX1006, [0143], [0160]. The combination of Kennewick and Hao therefore renders claim 10 obvious.

10. Claim 11

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

277. Kennewick in combination with Hao renders the method of claim 1 obvious.

[11.a] “invoking, by the one or more processors, one or more system resources to select the purchase opportunity based on the determined context;”

278. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the main speech processing unit (e.g., the “*one or more processors*”) “invok[es] the proper resources, including agents” such as a domain agent to process and respond to each received natural language utterance. EX1006, [0017], [0028]. These invoked resources including agents are “*one or more system resources,*” being resources that are available for use by the system. For example, FIG. 5 of Kennewick illustrates the speech recognition engine, parser, and agents are all included in the main speech processing unit 98, and are resources that can be invoked by the main speech processing unit 98.

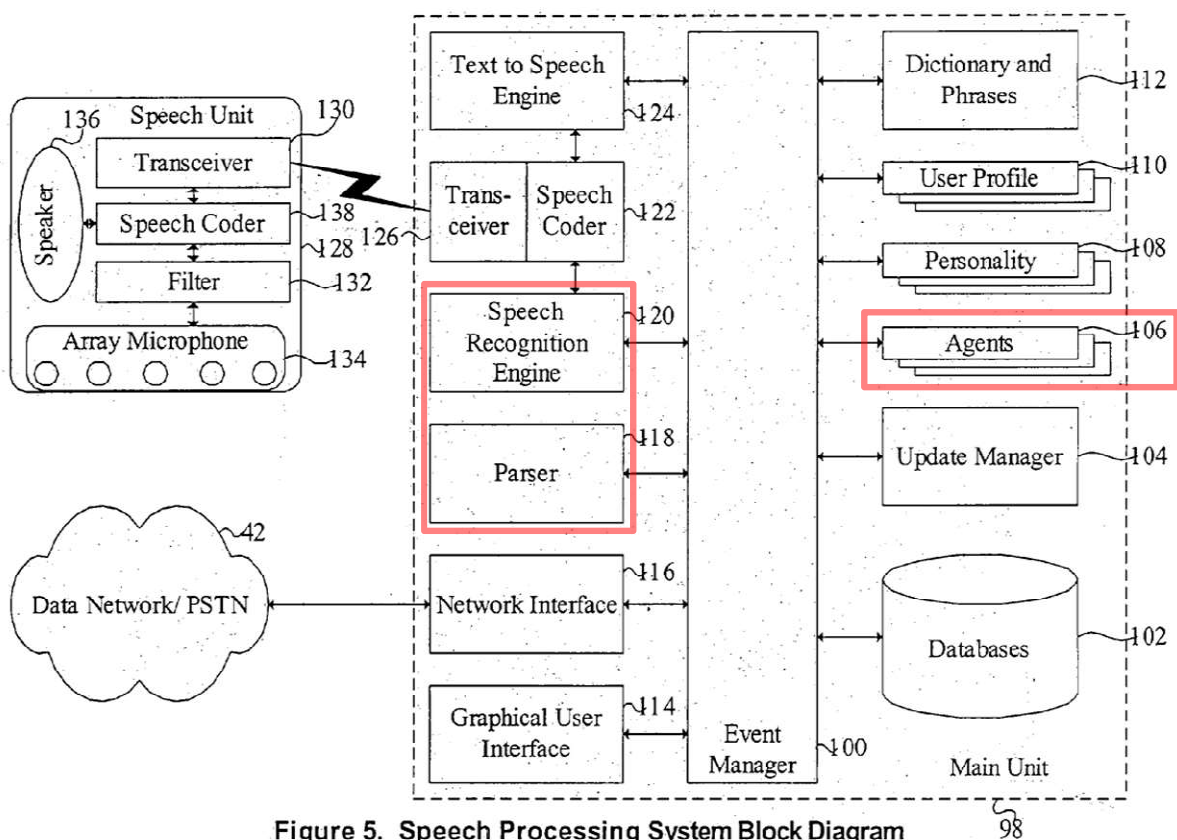


Figure 5. Speech Processing System Block Diagram

Id., FIG. 5.

279. In addition, Kennewick discloses invoking the one or more system resources *“to select the purchase opportunity based on the determined context.”* For example, Kennewick discloses “invok[ing] the correct agent” based on the determined context. EX1006, [0162]. A POSITA would have found it obvious that when the main speech processing unit determines that the context of the user’s utterance relates to an e-commerce context such as “remote ordering” or “fast food ordering,” it invokes the appropriate system resources including an appropriate

domain agent for the determined area of functionality, which it uses to select the purchase opportunity. *See also id.*, [0018], [0160] (“The context of a question or command may determine the domain and thereby, the domain agent 156, if any, to be invoked”). Thus, the invoked system resources select the purchase opportunity based on the determined context, due to being selected by the domain agent invoked based on the determined context.

280. To the extent that Kennewick does not explicitly disclose the invoked resources selecting the purchase opportunity based on the determined context, the combination with Hao renders it obvious. For example, Hao discloses analyzing the meaning of the user’s request “in an e-commerce context,” and, based on the determined meaning, selecting items with which to populate a “shopper specific shopping cart” to be presented to the user. EX1007, [0040], [0043]. As discussed above, a POSITA would have found it obvious to modify Kennewick’s system to implement Hao’s e-commerce functionality as part of one or more e-commerce-related domains in Kennewick’s system, e.g., as a domain for facilitating the remote ordering of goods or services. Thus, when the combined system determines that the user’s natural language utterance corresponds to an e-commerce/remote ordering context, Kennewick’s main speech processing unit invokes a domain agent that functions in accordance with the teachings of Hao to process the user’s request and select one or more items constituting a purchase opportunity. Thus, the combination

of Kennewick and Hao renders obvious “*invoking . . . one or more system resources to select the purchase opportunity based on the determined context,*” because the system resources include the domain agent selected based on the determined context, which selects the purchase opportunity based on the teachings of Hao. The combination of Kennewick and Hao therefore renders limitation [11.a] obvious.

[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.”

281. Kennewick in combination with Hao renders this limitation obvious. For example, as discussed above, Kennewick discloses that the invoked “*one or more system resources*” includes “agents,” such as a “domain agent” that provides a “convenient and re-distributable package[] or module[] for [a corresponding] application area.” EX1006, [0023], [0126]. Each domain agent in Kennewick 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 commands pertaining to its domain. *Id.*, claim 1, [0017]-[0018] (agents are “packages of executable code, scripts, links to information, data, and other data forms”), [0126]. The combination of Kennewick and Hao therefore renders limitation [11.b] obvious.

11. Claim 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.”

282. Kennewick in combination with Hao renders claim 12 obvious. First, Kennewick discloses that the “*agent*,” e.g., domain agent, “*returns results of user questions*” as “*responses to users*,” where the response is 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]. For example, the invoked agent is configured to “receive, process and respond to user questions, queries and commands” by “creat[ing] one or more queries to one or more local or external information sources” and “evaluat[ing] the results of the one or more queries” with which to construct “a satisfactory response” to the user. EX1006, [0017], [0167], [0170]-[0171], [0175]. In addition, the agent, when creating the response, “select[s] the presentation[/system] personality 108 to be used,” and “appl[ies] a number of transformations to the tokens before presentation to the user” that are “derived from a number of sources including . . . the user profile 110.” *Id.*, [0178], [0180]; *see also id.*, [0156] (“system personality 108”).

283. In the combination of Kennewick and Hao discussed above, a POSITA would have understood when the user makes a request in “an e-commerce context,” the system would invoke an appropriate agent as taught by Kennewick, which

queries an inventory data store for purchasable items matching the user's request as taught by Hao. *See* EX1007, [0030], [0032] (describing searching for “[a]vailable items . . . specified in an inventory data store 223 of an e-commerce server 222”), [0035]. A POSITA would have understood that in the combined system, the agent uses the results of the queries, corresponding to results of “*user questions, queries, and/or commands,*” to create responses to the user as taught by Kennewick, using “*system personality, and user preferences or other data in the user-specific profile*” in the same way as taught by Kennewick. The combination of Kennewick and Hao therefore renders claim 12 obvious.

12. Claim 13: “The method of claim 11, further comprising updating the agent from a remote location over a network to provide new functionality.”

284. Kennewick in combination with Hao renders claim 13 obvious. First, Kennewick discloses “*updating the agent from a remote location over a network.*” For example, Kennewick discloses an “update manager” that “manages the automatic and manual loading and updating of agents and their associated data from the Internet.” EX1006, [0120]; *see also id.*, [0017] (“agents and their associated data can be updated remotely over a network as new behavior is added or new information becomes available”); [0126] (“agents and their associated data can be updated remotely over a network as new behavior is added or new information

becomes available”). FIG. 5 reproduced below illustrates the “update manager 104” implemented as part of Kennewick’s main speech processing system 98.

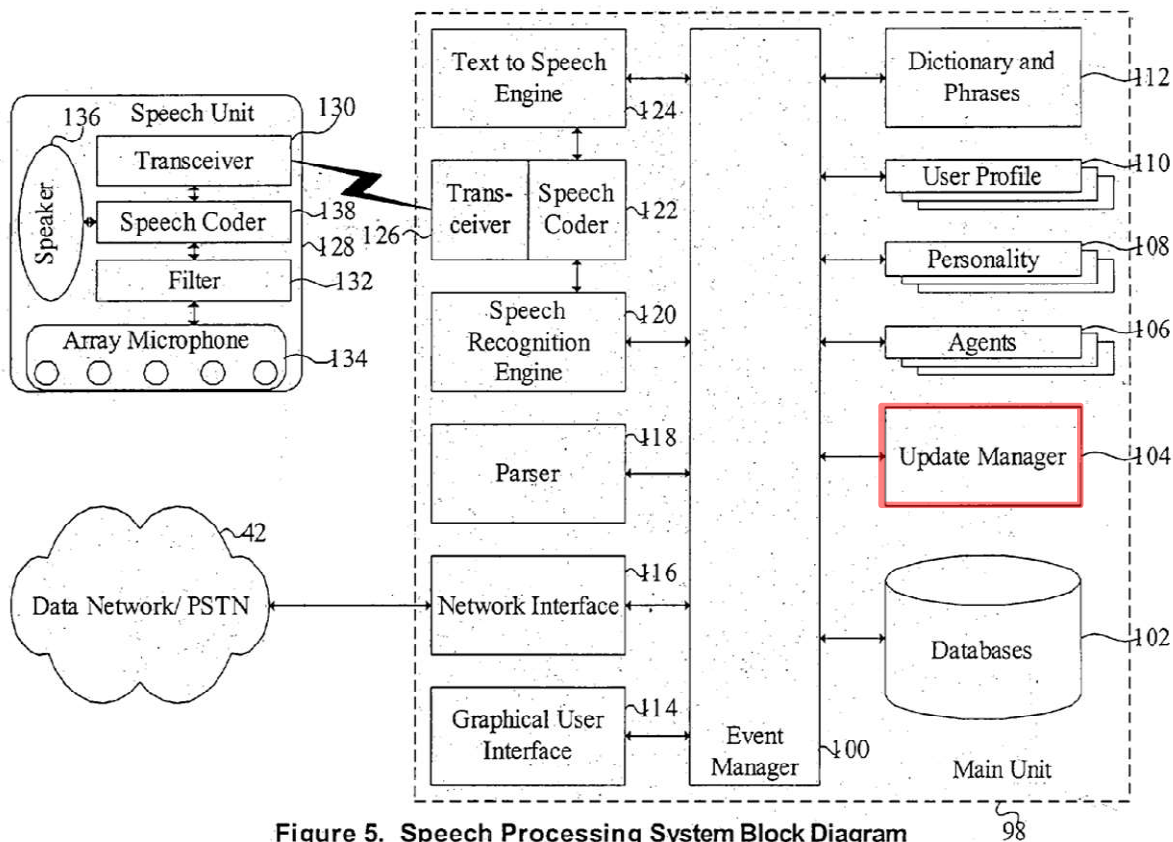


Figure 5. Speech Processing System Block Diagram

Id., FIG. 5. Thus, Kennewick discloses “**updating the agent from a remote location over a network,**” e.g., the Internet, via the Update Manager.

285. Kennewick further discloses updating the agent to “**provide new functionality.**” For example, Kennewick discloses updating agents to “add[] new behaviors,” which a POSITA would have understood may correspond to providing “**new functionality.**” EX1006, [0017], [0019], [0126]; *see also id.*, [0129] (“[d]evelopers of agents 106 can add new functionality to the agent library 158 as

required”). The combination of Kennewick and Hao therefore renders claim 13 obvious.

13. Claim 14: “The method of claim 11, wherein the agent uses system resources and the services of other agents.”

286. Kennewick in combination with Hao renders claim 14 obvious. For example, Kennewick discloses that the agent may use “system resources and the services of other, typically more specialized, agents.” EX1006, [0017].

14. Claim 15: “The method of claim 11, further comprising providing a framework allowing the user to create or modify the agent.”

287. Kennewick in combination with Hao renders claim 15 obvious. For example, Kennewick discloses providing a speech interface or graphical user interface through which the user “can create and manage modifications to agents.” EX1006, [0134]. The speech or graphical interface allowing the user to create or modify the agent meets the claimed “*framework allowing the user to create or modify the agent.*”

15. Claim 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.”

288. Kennewick in combination with Hao renders claim 16 obvious. For example, Kennewick discloses “*providing a license management framework*” because it discloses an “agent manager 154” that provides a “license management client” that provide “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] (available license management schemes include “out right purchase, subscription for updates, one time or limited time use”). *Id.* Thus, Kennewick discloses providing a license management framework to enable a third party to sell an agent, and for the user to purchase the agent, on a one time or subscription basis. The combination of Kennewick and Hao therefore renders claim 16 obvious.

16. Claim 17

[17.pre] “The method of claim 11, wherein:”

289. Kennewick in combination with Hao renders the method of claim 11 obvious.

[17.a] “the method is implemented on a network of devices using a common base of agents, data, information, user-specific profiles, and histories the user can interact with, and receive the same services and applications from, any one of the devices on the network; and”

290. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that its system that implements the claimed 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. The network of devices “can be placed at different locations throughout a home, place of business, vehicle or other location,” and the user can “*interact with, and receive the same services and applications* at any location equipped with the required device on the network.” EX1006, [0016]. Thus, Kennewick discloses that the user can interact with, and receive the same services and applications from *any*

one of the devices on the network. The combination of Kennewick and Hao therefore renders limitation [17.a] obvious.

[17.b] “the location of a particular device addressed by the user is used as part of context for questions asked of the user.”

291. Kennewick in combination with Hao renders this limitation obvious. For example, Kennewick discloses that the system may be deployed on “multiple devices” “placed at different locations,” where the system “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]. The combination of Kennewick and Hao therefore renders limitation [17.b] obvious.

17. Claim 18

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

292. Kennewick in combination with Hao renders the method of claim 1 obvious.

[18.a] “capturing, by the one or more processors, the first natural language utterance;”

293. Kennewick in combination with Hao renders this limitation obvious. The claimed term “the first natural language utterance” lacks antecedent basis. For my analysis, I assume “the first natural language utterance” in this limitation refers to the “natural language utterance” of limitation [1.a]. Kennewick discloses that the main speech processing unit (e.g., “*the one or more processors*”) includes “a speech recognition module for *capturing* the user input” received as “spoken natural

language queries, commands and/or other utterances from a user” at a speech unit. EX1006, [0013], [0014]; *see also id.*, [0028] (“capturing the user's question or query through accurate speech recognition”). A POSITA would have understood that the captured user input, e.g., “spoken natural language queries, commands and/or other utterances,” meets the claimed “***first natural language utterance.***” The combination of Kennewick and Hao therefore renders limitation [18.a] obvious.

[18.b] “parsing, by the one or more processors, the first natural language utterance;”

294. Kennewick in combination with Hao renders this limitation obvious. For example, Kennewick discloses that the main speech processing unit (e.g., “***the one or more processors***”) includes “a parser for ***parsing*** the input,” which, as I discussed above, includes the claimed “***first natural language utterance.***” EX1006, [0013]; *see also id.*, [0028] (“parsing and interpreting the question or query”). The combination of Kennewick and Hao therefore renders limitation [18.b] obvious.

[18.c] “interpreting, by the one or more processors, the first natural language utterance; and”

295. Kennewick in combination with Hao renders this limitation obvious. For example, Kennewick discloses that the parser included in the main speech processing unit (e.g., “***the one or more processors***”) performs “parsing and ***interpreting*** [of] the [user’s] question or query” corresponding to the claimed “***first natural language utterance.***” EX1006, [0028]. Thus, Kennewick discloses that the

main speech processing unit performs “interpreting” of the first natural language utterance. The combination of Kennewick and Hao therefore renders limitation [18.c] obvious.

[18.d] “invoking, by the one or more processors, an applicable agent from a library of agents.”

296. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the main speech processing unit (e.g., “***the one or more processors***”) includes the parser which, based on its interpretation of the “first natural language utterance,” “invoke[s] the required agent” to process and respond to the user’s request. EX1006, [0028], [0127]-[0128]; *see also id.*, [0032] (“the system can invoke the correct agent” based on the determined “most[] likely context or domain for a user’s question or command”). The invoked “required agent” meets the claimed “***applicable agent***,” being an agent that is applicable to the processing of the first natural language utterance. For example, the applicable agent may be a “domain agent 156” invoked based on the determined context of the utterance. *See id.*, [0160].

297. In addition, Kennewick discloses that the applicable invoked agent (e.g., a domain agent 156) is selected from “***a library of agents***.” For example, FIG. 6 of Kennewick illustrates an “agent architecture” that includes a collection or library of domain agents managed by an “agent manager.”

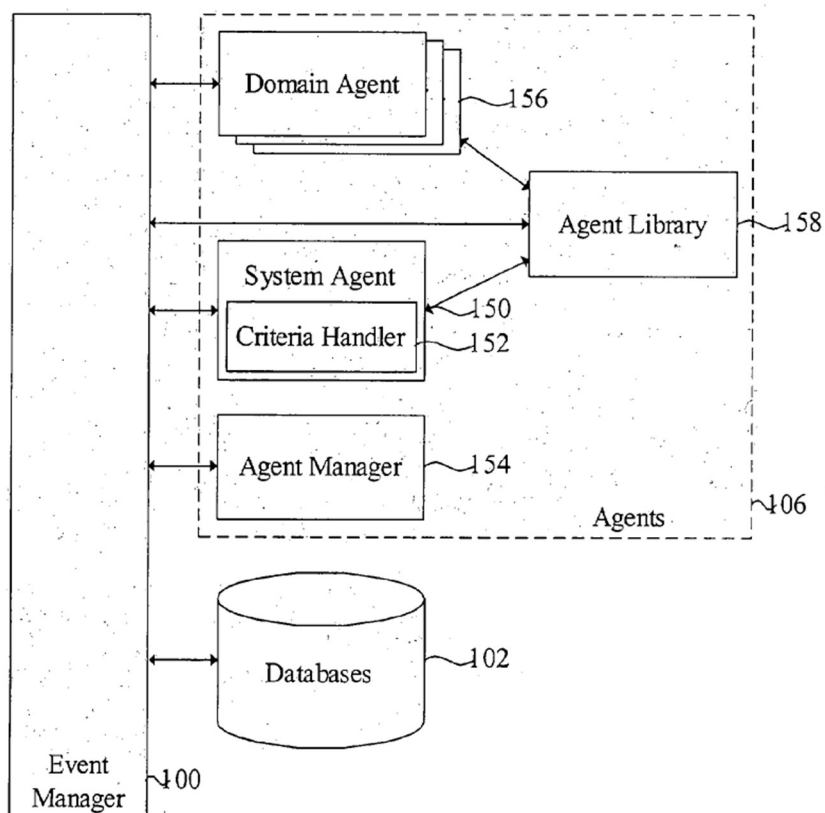


Figure 6. Agent Architecture

See EX1006, [0126], FIG. 6. A POSITA would have understood that a managed collection of agents may correspond to a “library” of agents. The combination of Kennewick and Hao therefore renders limitation [18.d] obvious.

18. Claim 19

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

298. Kennewick in combination with Hao renders the method of claim 1 obvious.

[19.a] “invoking an agent to process the natural language utterance; and”

299. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the main speech processing unit “determin[es] the domain of expertise required and context” of the natural language utterance, and based on the determined context, “invoke[s] the required agent.” EX1006, [0028], [0032], [0127]-[0128], [0160]. For example, the parser processes the “recognized words and phrases” of the natural language utterance, and “transforms them into complete commands and questions using data supplied by the agents.” *Id.*, [0123]. The invoked agent “then ***process[es]*** the commands or questions” corresponding to the natural language utterance. *Id.*, [0123]; *see also id.*, [0017] (“Agents are executables that receive, process and respond to user questions, queries and commands”). Thus, Kennewick discloses that the claimed “***invoking an agent to process the natural language utterance,***” because the invoked agent processes the question or command of the natural language utterance. *See also id.*, [0165] (“the parser can evoke the correct agent to process the question or command”). The combination of Kennewick and Hao therefore renders limitation [19.a] obvious.

[19.b] “creating a query to a local database or through a network interface to a remote database.”

300. Kennewick in combination with Hao renders this limitation obvious. Kennewick discloses that the invoked agent “process[es] the commands or questions” of the received natural language utterance by “create[ing] queries to local

databases 102 or th[r]ough [a] network interface 116 to data sources on the Internet 42 or other networks.” EX1006, [0123]. Thus, Kennewick discloses the agent performs “*creating a query to a local database or through a network interface to a remote database*,” e.g., a data source “on the Internet.” The combination of Kennewick and Hao therefore renders limitation [19.b] obvious.

19. Claim 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.”

301. Kennewick in combination with Hao renders claim 20 obvious. Kennewick discloses that the user can use a user interface such as a “natural language interactive speech interface” to “modify existing questions and queries” and “set preferences and specify presentation parameters for results.” EX1006, [0021], [0073]. Thus, Kennewick discloses providing a “*user interface*” (e.g., the “natural language interactive speech interface”) “*allowing the user to modify questions, queries*,” or “*preferences, or parameters*” (e.g., user preferences and “presentation parameters”). In addition, Kennewick further discloses providing a user interface allowing the user to modify the user’s user-specific profile, by “manually enter[ing]” user profile data, which may include “user specific parameters.” *Id.*, [0142]-[0143]; *see also id.*, [0125] (“graphical user interface” can “be used as a substitute or complement to the speech interface”). The combination of Kennewick and Hao therefore renders claim 20 obvious.

20. Claim 21: “The method of claim 1, wherein the context comprises a categorization of the recognized words or phrases.”

302. Kennewick in combination with Hao renders claim 21 obvious. As discussed above for limitation [1.c], Kennewick discloses determining “a context for an utterance” based on the recognized words or phrases through “keyword matching” to “determine the domain” of the utterance. EX1006, [0160]. A POSITA would have understood that “keyword matching” involves matching recognized words or phrases of the utterances with keywords associated with the context. *Id.* Kennewick further discloses that the contexts or domains “are broadly divided into two categories, query and response, and control.” EX1006, [0018]. Thus, Kennewick discloses that the determined context comprises “*a categorization of the recognized words or phrases,*” because it matches the recognized words and phrases to the context or domain determined for the utterance, which categorizes them as being associated with “query and response” or “control.” The combination of Kennewick and Hao therefore renders claim 21 obvious.

21. Claim 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.”

303. Kennewick in combination with Hao renders claim 22 obvious. As discussed above, Kennewick discloses determining the “context or domain for a user’s question or command,” where the possible contexts and domains include

query and response domains pertaining to “driving directions, travel services, entertainment scheduling, and other information,” and/or control domains for “control of specific devices.” EX1006, [0018], [0132], [0161]. Other contexts include “[c]ontrol of vehicle multimedia entertainment electronics,” “[l]ocation based marketing,” “remote ordering and payment,” “directories of travel and entertainment services,” “[w]eather information,” “news information,” “[s]ports scores,” “[i]nteractive educational programs,” subscriptions “to multimedia information channels,” etc. *Id.*, [0043]-[0051], [0060]-[0072], [0074]-[0085]. Thus, Kennewick discloses that the context comprises one of **commands** (e.g., control domain), **directories** (“directories of travel and entertainment services”), **e-commerce** (“remote 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. The combination of Kennewick and Hao therefore renders claim 22 obvious.

22. Claim 23

304. Claim 23 is rendered obvious for similar reasons as claim 1. The claimed “**system**” is met by Kennewick’s “interactive natural language speech system” that includes a “speech unit” and “main speech processing unit.” As

discussed above, these components perform “*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 program instructions, which when executed cause the one or more physical processors to*” perform certain steps. See EX1006, [0012], [0095]-[0097], [0111], [0115], claim 1. The remaining elements of claim 23 are substantively identical to those of claim 1 and are obvious for the same reasons.

23. Claim 24

305. Claim 24 is substantively identical to claim 2 and is obvious for the same reasons.

24. Claim 26

306. Claim 26 is substantively identical to claim 6 and is obvious for the same reasons.

25. Claim 27

307. Claim 27 is substantively identical to claim 5 and is obvious for the same reasons.

26. Claim 28

308. Claim 28 is substantively identical to claim 6 and is obvious for the same reasons.

27. Claim 29

309. Claim 29 is substantively identical to claim 7 and is obvious for the same reasons.

28. Claim 31

310. Claim 31 is substantively identical to claim 9 and is obvious for the same reasons.

29. Claim 32

311. Claim 32 is substantively identical to claim 10 and is obvious for the same reasons.

30. Claim 33

312. Claim 33 is substantively identical to claim 11 and is obvious for the same reasons.

31. Claim 34

313. Claim 34 is substantively identical to claim 12 and is obvious for the same reasons.

32. Claim 35

314. Claim 35 is substantively identical to claim 13 and is obvious for the same reasons.

33. Claim 36

315. Claim 36 is substantively identical to claim 14 and is obvious for the same reasons.

34. Claim 37

316. Claim 37 is substantively identical to claim 15 and is obvious for the same reasons.

35. Claim 38

317. Claim 38 is substantively identical to claim 16 and is obvious for the same reasons.

36. Claim 39

318. Claim 39 is substantively identical to claim 17 and is obvious for the same reasons.

37. Claim 40

319. Claim 40 is substantively identical to claim 18 and is obvious for the same reasons.

38. Claim 41

320. Claim 41 is substantively identical to claim 19 and is obvious for the same reasons.

39. Claim 42

321. Claim 42 is substantively identical to claim 20 and is obvious for the same reasons.

40. Claim 43

322. Claim 43 is substantively identical to claim 21 and is obvious for the same reasons.

41. Claim 44

323. Claim 44 is substantively identical to claim 22 and is obvious for the same reasons.

E. Ground 5: Kennewick in combination with Hao and Aretoulaki render obvious claims 3 and 25

1. Claim 3

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

324. Kennewick in combination with Hao and Aretoulaki renders the preamble obvious. As discussed above, Kennewick in combination with Hao renders obvious the method of claim 1.

[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;”

325. Kennewick in combination with Hao and Aretoulaki renders this limitation obvious. For example, as discussed for limitation [1.e], Kennewick in combination with Hao discloses “delivering . . . the selected purchase opportunity.” Hao further discloses that after the purchase opportunity is delivered to the user, the user can select from a number of “shopping interactive options” through “a natural language input element.” EX1007, [0011]. These include options to effect “at least one of a change in the items presented in the shopping cart,” or the “finaliz[ing] of an e-commerce transaction involving a purchase of the items presented in the shopping cart.” *Id.* Hao’s input processor (or in the combined system, Kennewick’s

main speech processing unit modified to incorporate the functionality of Hao's input processor for selecting and interacting with purchase opportunities), which corresponds to the claimed "*one or more processors,*" receives "*an indication of*" the user exercising one of these interactive options, which corresponds to an "*occurrence of a user interaction with the purchase opportunity.*" The combination of Kennewick and Hao thus discloses "*receiving . . . an indication of an occurrence of a user interaction with the purchase opportunity,*" e.g., the user selecting an option to change the items in the presented cart or purchase the items in the cart.

326. In addition, a POSITA would have further been motivated to modify the combination of Kennewick and Hao based on the teachings of Aretoulaki to receive "*an indication of an occurrence of a user interaction with the purchase opportunity*" included in the interaction pattern. Aretoulaki discloses a system (a "knowledge communications platform") for providing "intuitive natural language (text or speech) dialogue-based electronic commercial transactions and information exchanges." EX1004, Abstract. For example, Aretoulaki's system analyzes received user inputs to determine a "Result Set" of products or services which it presents to the user as a purchase opportunity. *Id.*, 8:27-46, 9:12-20, 13:37-40, 15:44-51.

327. As discussed above, Aretoulaki discloses tracking user interactions with a presented purchase opportunity, where the user can “correct the system, or to interrupt it in the case of a misunderstanding.” EX1004, 7:37-41. By “constantly monitoring” the system’s performance with regards to “user behavior,” the system can “locate communication or understanding problems” within the dialogue, e.g., “[b]reaks in the normal flow, such as . . . objections on the part of the user or erroneous recognition on the part of the system,” and in response “trigger a repair mechanism” to correct the misunderstanding. *Id.*, 7:10-24, 18:19-25; *see also id.*, 32:21-25 (monitoring user interactions to identify “a correction of [a] parameter value”). 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” or tracking the user’s interactions with the system, including interactions with any presented purchase opportunities.

328. In my opinion, a POSITA would have been motivated to combine Aretoulaki with Kennewick and Hao. The references are analogous, all relating to speech-based interfaces through which a user can use natural language utterances to search for and purchase products. *See, e.g.*, EX1006, Abstract, [0018], [0066]-[0068]; EX1007, Abstract, [0009], [0011], [0040]; EX1004, Abstract, 8:27-48, 11:31-36, 22:10-23. These references also all describe tracking user interactions

with the system, including interactions with presented purchase opportunities, to determine user interests and preferences. *See, e.g.,* EX1006, [0007] (“some questions can only be interpreted in the context of . . . the user's history of interests and preferences”), [0067] (“The system facilitates remote ordering by using . . . customer preferences, customer order history, etc.”), [0143] (“[p]ersonal profiles may include information specific to the individual, their interests”); EX1007, [0027] (“[t]he disambiguation engine 138 can utilize data, rules, and/or preferences stored in data store 140 to resolve ambiguities”; shopper profiles may include “customer preferences”), [0033] (“Ambiguities can be resol[v]ed and/or accuracy of other entries improved by using shopper specific information maintained within a shopper profile/preferences data store 228.”), [0046] (describing “resolv[ing] ambiguities using purchase history, preference, and profile information of the shopper”); EX1004, 9:47-49 (“The user model contains data about the general preferences and assumptions of individual users,” based on “monitored” user behavior).

329. A POSITA in possession of Kennewick, Hao, and Aretoulaki would therefore have been motivated to combine them, e.g., by modifying the combined system of Kennewick and Hao to, after presenting a purchase opportunity to the user, “constantly monitor” subsequent user interactions with the system “to locate [any potential] communication or understanding problems” that may occur. *See* EX1004, 7:11-35. The modification would give the system additional flexibility when

conducting a dialogue with the user, by enabling the user “to correct the system, or to interrupt it in the case of a misunderstanding” so that the system can recover from misunderstandings by executing an appropriate repair strategy. *Id.*, 7:39-41. A POSITA would further have been motivated to make the modification as it aligns with Kennewick’s stated goal of creating a system that can accommodate “full or partial failure” and achieve “graceful recovery.” EX1006, [0011].

330. The modification is a combination of prior art elements disclosed in Kennewick, Hao, and Aretoulaki according to known methods to yield predictable results. For example, Hao discloses that the user can take a number of different actions following presentation of a purchase opportunity, such as to “edit, update, and otherwise modify entries of the shopping cart,” or to finalize an “e-commerce transaction” to purchase the item. EX1007, [0011], [0034]. Aretoulaki further discloses flexible ways in which the user can interact with a purchase opportunity, where the user can “correct the system, or [] interrupt it in the case of a misunderstanding,” and the system, responsive to receiving indications of these interactions, responds accordingly to “dynamically establish the next step in the dialogue.” EX1004, 7:11-41. A POSITA would have had a reasonable expectation of success because the combination is a simple combination of known techniques for dialogue management and how the system “determin[es] the goal and the content of the next system action and message” based on received user interactions. *See*,

e.g., EX1004, 9:22-35; EX1006, [0118] (describing making “maximum use of context [and] prior information”). A POSITA would have found this modification obvious as a combination of prior art elements according to known methods, to create a more robust and natural interface for facilitating remote ordering of goods and services that can flexibly respond to different types of user interactions and recover from misunderstandings. The combination of Kennewick, Hao, and Aretoulaki therefore renders limitation [3.a] obvious.

[3.b] “determining, by the one or more processors, whether the interpretation of the natural language utterance was correct based on the indication; and”

331. Kennewick in combination with Hao and Aretoulaki discloses this limitation. To the extent that Kennewick and Hao do not disclose determining “*whether the interpretation of the natural language utterance was correct based on the indication*” of an occurrence of a user interaction with the purchase opportunity, the combination with Aretoulaki renders it obvious.

332. As discussed above, Aretoulaki discloses determining “*whether the interpretation of the natural language utterance was correct based on the indication*” of an occurrence of a user interaction with the purchase opportunity, by determining whether the indication of the user’s interaction 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” (i.e., the claimed “*natural language utterance*”). EX1004, 12:30-35, 38:44-51.

333. In the combined system, a POSITA would have found it obvious to modify Kennewick and Hao to implement Aretoulaki’s repair mechanism to correct misunderstandings that occur during a conversation. For example, Kennewick teaches that the system makes “significant use of context [and] prior information” to interpret user utterances. EX1006, Abstract, [0011], [0118]. It would be obvious to use context and prior information to determine whether a user interaction represents a “break[] in the normal flow” of conversation that signifies a misunderstanding as taught in Aretoulaki, to enable the system to detect and repair the misunderstanding. *See* EX1004, 18:8-25 (describing monitoring “the evolution of the dialogue” and prior information to detect when to “trigger a repair mechanism”). This would allow for a more reliable and robust natural language interface that is able to locate potential “communication or understanding problems” during a conversation, and “dynamically establish the next step in the dialogue” to react to and repair the misunderstanding, enabling the system that can accommodate “full or partial failure” and achieve “graceful recovery.” EX1004, 7:11-35; EX1006, [0011].

334. The modification is a combination of prior art elements disclosed in Kennewick, Hao, and Aretoulaki according to known methods to yield predictable results. POSITA would have had a reasonable expectation of success because the

modification is a combination of known prior art techniques for dialogue management and “determining the goal and the content of the next system action and message” based on received user interactions (e.g., Kennewick’s use of context and prior information to interpret utterances, and Aretoulaki’s use of tracked interactions and prior information to determine when the system has misunderstood the user’s request), 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 misunderstood the user’s request). *See, e.g.,* EX1006, [0118]; EX1004, 9:22-35, 18:8-25. The combination of Kennewick, Hao, and Aretoulaki therefore renders limitation [3.b] obvious.

[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.”

335. Kennewick in combination with Hao and Aretoulaki renders this limitation obvious. For example, Kennewick discloses that the main speech processing unit (e.g., “***one or more processors***”) includes a parser that interprets natural language utterances using “context [and] prior information.” EX1006, Abstract, [0011]. Thus, a POSITA would have found it obvious that the system would interpret a “***subsequent natural language utterance***” based on prior information, which as discussed above includes information of “***whether the interpretation of the [previous] natural language utterance was correct.***”

336. In addition, to the extent that Kennewick and Hao do not disclose *“interpreting . . . a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct,”* the combination with Aretoulaki renders it obvious. As discussed above, Aretoulaki discloses triggering a “repair mechanism” when it determines that its (previous) interpretation of the utterance was incorrect, e.g., by determining that the user’s interaction with the purchase opportunity corresponds to a “Correct” dialog act. EX1004, 18:19-25, 38:44-39:6. The triggered repair mechanism updates the system’s interpretation of the natural language utterance by establishing a “new/modified parameter value.” EX1004, 15:37-42, 18:19-25, 39:1-5. Afterwards, the system interprets a subsequent natural language utterance based on the determination of whether the interpretation of the natural language utterance was correct, by interpreting the utterance “in terms of the preceding exchanges between the user and the system,” including the “new/modified parameter value” set by the repair mechanism. EX1004, 4:17-23, 39:1-5; *see also id.*, 16:38-40 (“The generated expectations about the next user input and its meaning in the context of what has taken place before guide the interpretation of the [current] input”), 22:28-31 (“[t]he system always keeps a record of all the exchanges in a specific interaction session with the user (discourse history) . . . to contextually interpret each new user input”). Thus, the subsequent utterances are *“interpret[ed] . . . based on the determination*

of whether the interpretation of the [prior] natural language utterance was correct.” This is consistent with the disclosure of the ’758 patent, which explains “previous utterances” determined to have been interpreted incorrectly “may be reinterpreted to improve subsequent interpretations”. EX1001, 5:28-32.

337. A POSITA would have understood that in the combined system discussed above, in which the Kennewick-Hao combination is further modified based on 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 would interpret the subsequent natural language utterance in the same way as taught by Aretoulaki. A POSITA would have found this combination obvious, as it would allow for a more reliable and robust natural language interface that is able to recover from detected misunderstandings, and because the combination aligns with Kennewick’s teachings of making “significant use of context [and] prior information” when interpreting received utterances. EX1006, Abstract, [0011]. 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. The combination of Kennewick, Hao, and Aretoulaki therefore renders limitation [3.c] obvious.

2. Claim 25

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

F. Ground 6: Kennewick in combination with Hao and Jong render obvious claims 8 and 30

1. Claim 8: “The method of claim 1, wherein recognizing the words or phrases from the natural language utterance comprises: 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”

339. Kennewick in combination with Hao and Jong renders claim 8 obvious. Kennewick and Hao each disclose “*recognizing the words or phrases from [a] natural language utterance.*” For example, as discussed above, Kennewick discloses a “speech recognition engine” recognizing the words or phrases from the natural language utterance. *See, e.g.,* EX1006, [0123], [0155]. Similarly, Hao also discloses “an automatic speech recognition (ASR) engine” that processes “speech input” to generate text meeting the claimed “words or phrases.” EX1007, [0022], claims 8, 15.

340. To the extent that Kennewick and Hao do not explicitly disclose the speech recognition engine using “*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,” the combination with Jong renders it obvious. As discussed above, Jong’s “speech recognition device” performs speech recognition by translating “speech input signals into textual representations . . . representing” a “*stream of phonemes.*” EX1008, Abstract, 9:34-40. The speech recognition device detects the phonemes using a “word model device” that “includes a listing of phonemes and corresponding textual code representations of these phonemes,” which a POSITA would have understood corresponds to using “*phonetic dictation to recognize [the] stream of phonemes in the natural language utterance.*” *Id.*, 9:41-46. Jong further discloses that the “*recognized words or phrases [of the natural language utterance] are based on the recognized stream of phonemes*” because it uses a “subword model database 304 contain[ing] word syllables that are correlated with the phonemes of the word model device 303” to extract “corresponding syllables” from the phonemes, and checks the determined syllables “against the lexicon data stored in [a] lexicon database” to group the syllables into “recognizable words.” EX1008, 5:50-67. In other words, the “recognizable words” of the speech signal (“*recognized words or phrases*”) are based on the recognized stream of phonemes, being grouped from syllables identified from the recognized phonemes. *See also id.*, 10:57-62 (claim 8, “determining phonemes included in the first speech input signals and converting the phonemes into . . . textual words”).

341. Kennewick, Hao, and Jong are analogous, all describing systems that perform speech recognition. *See, e.g.*, EX1006, [0123] (describing speech recognition engine using “dictionary and phrases module 112”); EX1007, [0022] (“speech input can be processed by an automatic speech recognition (ASR) engine”); EX1008, 5:50-67 (converting speech signals to phonemes to syllables to words using subword models and a stored lexicon). While Kennewick-Hao leaves open the specific technique it uses to recognize speech, Jong teaches a specific speech recognition technique that a POSITA would have recognized as straightforward to implement in Kennewick-Hao’s system. A POSITA would have been motivated to implement the speech recognition engine of Kennewick-Hao to perform phonetic dictation to recognize a stream of phonemes in the natural language utterance, and recognize words or phrases based on the recognized stream of phonemes, as taught by the speech recognition device of Jong, as a specific way for the system to perform speech recognition to recognize words and phrases from a user’s voice input.

342. The modification is a combination of prior art elements disclosed in Kennewick, Hao, and Jong according to known methods to yield predictable results. A POSITA would have had a reasonable expectation of success because the combination is a simple substitution of a specific known technique for speech recognition (e.g., Jong’s phonetic dictation that recognizes phonemes in a natural language utterance, and recognizes words based on the phonemes) for more general

disclosure of speech recognition found in Kennewick and Hao, to yield predictable results, e.g., speech recognition performed using Jong's specific technique. A POSITA would have also found the modification obvious to try as there were a finite number of well-known ways to perform speech recognition, and would have had a reasonable expectation of success that the speech recognition performed in Kennewick and/or Hao could be done using phonetic dictation to recognize phonemes as taught in Jong. The combination of Kennewick, Hao, and Jong therefore renders claim 8 obvious.

2. Claim 30

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

X. CONCLUSION

344. For the reasons I have presented above, it is my opinion that the references serving as the bases for the grounds of rejection in the Petition render claims 1-44 of the challenged patent unpatentable.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false

statements and the like so made are punishable by fine or imprisonment, or both,
under Section 1001 of Title 18 of the United States Code.

Dated: 7/1/2025

Padraic Smyth
Padhraic Smyth