

**UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

SAMSUNG ELECTRONICS CO. LTD. and SAMSUNG ELECTRONICS  
AMERICA, INC.,  
Petitioners,

v.

VB ASSETS, LLC,  
Patent Owner

---

IPR2025-00867

U.S. Patent No. 7,818,176

---

**DECLARATION OF STUART LIPOFF IN SUPPORT OF PETITION FOR  
*INTER PARTES* REVIEW OF U.S. PATENT NO. 7,818,176**

Mail Stop PATENT BOARD  
Patent Trial and Appeal Board  
U.S. Patent & Trademark Office  
P.O. Box 1450  
Alexandria, VA 22313-1450

## TABLE OF CONTENTS

|      |                                                                                                            |    |
|------|------------------------------------------------------------------------------------------------------------|----|
| I.   | Qualifications.....                                                                                        | 3  |
| II.  | Understanding of Relevant Legal Principles.....                                                            | 9  |
| III. | Technical Background.....                                                                                  | 11 |
| A.   | Language and Dialog .....                                                                                  | 11 |
| B.   | Interactive/Spoken Dialog Systems .....                                                                    | 13 |
| 1.   | Speech Recognition Component.....                                                                          | 17 |
| 2.   | Language Understanding .....                                                                               | 19 |
| 3.   | Dialog Management Component .....                                                                          | 21 |
| 4.   | Response Generator and Speech Synthesizer .....                                                            | 22 |
| C.   | Intent-Based Advertising.....                                                                              | 23 |
| D.   | Web-Based Advertising Tools .....                                                                          | 25 |
| 1.   | Early Web-Advertising Systems.....                                                                         | 26 |
| 2.   | User Modelling Approaches .....                                                                            | 29 |
| 3.   | Using Analytics to Improve Models.....                                                                     | 33 |
| IV.  | '176 Patent .....                                                                                          | 35 |
| A.   | Overview of the '176 Patent.....                                                                           | 35 |
| B.   | Level of Ordinary Skill in the Art.....                                                                    | 45 |
| C.   | Prosecution and Litigation History .....                                                                   | 46 |
| 1.   | U.S. Prosecution Before the U.S. Patent & Trademark<br>Office.....                                         | 46 |
| 2.   | Inter Partes Review before the Patent Trial and Appeal<br>Board.....                                       | 49 |
| 3.   | Proceedings in the U.S. District Court.....                                                                | 50 |
| D.   | Claim Construction .....                                                                                   | 51 |
| V.   | GROUND 1: THE COMBINATION OF TASCHEREAU AND<br>HANSEN RENDERS OBVIOUS CLAIMS 1-2, 9-11, 27-28, 35-37 ..... | 52 |
| A.   | Overview of the Combination .....                                                                          | 52 |
| 1.   | Taschereau .....                                                                                           | 53 |

## TABLE OF CONTENTS

(continued)

Page ii

|     |                                                                                                                               |     |
|-----|-------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.  | Hansen.....                                                                                                                   | 58  |
| 3.  | Motivation to Combine .....                                                                                                   | 66  |
| B.  | Independent Claims 1 and 27 .....                                                                                             | 70  |
| 1.  | Preamble [1P] / Preamble [27P] .....                                                                                          | 72  |
| 2.  | Receiving Natural Language Request [1A] / [27A] .....                                                                         | 73  |
| 3.  | Recognizing One or More Words in the Utterance [1B] /<br>[27B] .....                                                          | 77  |
| 4.  | Mapping Phonemes [1C] / [27C] .....                                                                                           | 81  |
| 5.  | Generating Preliminary Interpretation of Utterance [1D] /<br>[27D] .....                                                      | 88  |
| 6.  | Interpreting the Utterance, Including Establishing<br>Context [1E] / [27E].....                                               | 93  |
| 7.  | Selecting An Advertisement [1F] / [27F] .....                                                                                 | 100 |
| 8.  | Presenting the Selected Advertisement [1G] / [27G] .....                                                                      | 102 |
| C.  | Dependent Claims 2 and 28 .....                                                                                               | 104 |
| D.  | Dependent Claims 9 and 35 .....                                                                                               | 108 |
| E.  | Dependent Claims 10 and 36 .....                                                                                              | 110 |
| F.  | Dependent Claims 11 and 37 .....                                                                                              | 114 |
| VI. | GROUND 2: THE COMBINATION OF TASCHEREAU, HANSEN<br>AND FRANCO RENDERS OBVIOUS CLAIMS 12-18, 25-26, 38-<br>44, AND 51-52 ..... | 115 |
| A.  | Overview of the Combination .....                                                                                             | 115 |
| 1.  | Franco.....                                                                                                                   | 116 |
| 2.  | Motivation to Combine .....                                                                                                   | 119 |
| B.  | Independent Claims 14 and 40 .....                                                                                            | 120 |
| 1.  | Preamble [14P] / [40P].....                                                                                                   | 122 |
| 2.  | Receiving Natural Language Request [14A] / [40A] .....                                                                        | 123 |
| 3.  | Recognizing One or More Words in the Utterance [14B] /<br>[40B] .....                                                         | 123 |

# TABLE OF CONTENTS

(continued)

Page iii

|      |                                                                                                                          |     |
|------|--------------------------------------------------------------------------------------------------------------------------|-----|
| 4.   | Interpreting the Utterance Using Context [14C] / [40C] .....                                                             | 124 |
| 5.   | Selecting An Advertisement [14D] / [40D] .....                                                                           | 125 |
| 6.   | Presenting the Selected Advertisement [14E] / [40E] .....                                                                | 125 |
| 7.   | Determining An Incorrect Interpretation Occurred and<br>Reinterpreting [14F] / [40F] .....                               | 126 |
| C.   | Dependent Claims 12 and 38 .....                                                                                         | 134 |
| D.   | Dependent Claims 13, 15, 39 and 41 .....                                                                                 | 135 |
| E.   | Dependent Claims 16 and 42 .....                                                                                         | 140 |
| F.   | Dependent Claims 17, 18, 25, 26, 43, 44, 51, 52 .....                                                                    | 140 |
| VII. | GROUND 3: THE COMBINATION OF TASCHEREAU, HANSEN<br>AND JONES RENDERS OBVIOUS CLAIMS 3-8, 19-24, 29-34<br>AND 45-50 ..... | 140 |
| A.   | Overview of the Combination .....                                                                                        | 140 |
| 1.   | Jones .....                                                                                                              | 141 |
| 2.   | Motivation to Combine .....                                                                                              | 145 |
| B.   | Dependent Claims 3, 19, 29 and 45 .....                                                                                  | 148 |
| 1.   | Preamble [3P], [19P], [29P] and [45P] .....                                                                              | 148 |
| 2.   | Tracking an Interaction Pattern [3A], [19A], [29A] and<br>[45A] .....                                                    | 149 |
| 3.   | Updating a Model Based on the Interaction Pattern [3B],<br>[19B], [29B] and [45B] .....                                  | 151 |
| C.   | Dependent Claims 4, 20, 30 and 46 .....                                                                                  | 154 |
| D.   | Dependent Claims 5, 21, 31 and 47 .....                                                                                  | 158 |
| E.   | Dependent Claims 6, 22, 32 and 48 .....                                                                                  | 160 |
| 1.   | Preamble [6P], [22P], [32P] and [48P] .....                                                                              | 160 |
| 2.   | Building Long-Term and Short-Term Knowledge [6A],<br>[22A], [32A] and [48A] .....                                        | 161 |
| 3.   | Receiving a Subsequent Natural Language Utterance<br>[6B] and [22B] .....                                                | 165 |

# TABLE OF CONTENTS

(continued)

Page iv

|       |                                                                                          |     |
|-------|------------------------------------------------------------------------------------------|-----|
| 4.    | Interpreting Using Long-Term and Short-Term Knowledge [6C], [22C], [32B] and [48B] ..... | 166 |
| F.    | Dependent Claims 7, 23, 33 and 49 .....                                                  | 168 |
| G.    | Dependent Claims 8, 24, 34 and 50 .....                                                  | 171 |
| VIII. | GROUND 4: THE COMBINATION OF RAMER AND CROSS<br>RENDERS OBVIOUS CLAIMS 1- 52.....        | 172 |
| A.    | Overview of the Combination .....                                                        | 172 |
| 1.    | Ramer .....                                                                              | 172 |
| 2.    | Cross.....                                                                               | 179 |
| 3.    | Motivation to Combine .....                                                              | 183 |
| B.    | Independent Claims 1 and 27 .....                                                        | 186 |
| 1.    | Preamble [1P] / Preamble [27P] .....                                                     | 188 |
| 2.    | Receiving Natural Language Request [1A] / [27A] .....                                    | 191 |
| 3.    | Recognizing One or More Words in the Utterance [1B] /<br>[27B] .....                     | 193 |
| 4.    | Mapping Phonemes [1C] / [27C] .....                                                      | 196 |
| 5.    | Generating Preliminary Interpretation of Utterance [1D] /<br>[27D] .....                 | 199 |
| 6.    | Interpreting the Utterance Using Context [1E] / [27E] .....                              | 201 |
| 7.    | Selecting An Advertisement [1F] / [27F] .....                                            | 206 |
| 8.    | Presenting the Selected Advertisement [1G] / [27G] .....                                 | 209 |
| C.    | Independent Claims 14 and 40 .....                                                       | 210 |
| 1.    | Preamble [14P] / [40P].....                                                              | 212 |
| 2.    | Receiving Natural Language Request [14A] / [40A] .....                                   | 213 |
| 3.    | Recognizing One or More Words in the Utterance [14B] /<br>[40B] .....                    | 213 |
| 4.    | Interpreting the Utterance Using Context [14C] / [40C] .....                             | 214 |
| 5.    | Selecting An Advertisement [14D] / [40D] .....                                           | 215 |
| 6.    | Presenting the Selected Advertisement [14E] / [40E] .....                                | 215 |

# TABLE OF CONTENTS

(continued)

Page v

|    |                                                                                          |     |
|----|------------------------------------------------------------------------------------------|-----|
| 7. | Determining An Incorrect Interpretation Occurred and Reinterpreting [14F] / [40F].....   | 216 |
| D. | Dependent Claims 2, 18, 28, and 44 .....                                                 | 220 |
| E. | Dependent Claims 3, 19, 29 and 45 .....                                                  | 225 |
| 1. | Preamble [3P], [19P], [29P] and [45P] .....                                              | 226 |
| 2. | Tracking an Interaction Pattern [3A], [19A], [29A] and [45A] .....                       | 226 |
| 3. | Updating a Model Based on the Interaction Pattern [3B], [19B], [29B] and [45B].....      | 229 |
| F. | Dependent Claims 4, 20, 30 and 46 .....                                                  | 231 |
| G. | Dependent Claims 5, 21, 31 and 47 .....                                                  | 234 |
| H. | Dependent Claims 6, 22, 32 and 48 .....                                                  | 240 |
| 1. | Preamble [6P], [22P], [32P] and [48P] .....                                              | 240 |
| 2. | Building Long-Term and Short-Term Knowledge [6A], [22A], [32A] and [48A] .....           | 240 |
| 3. | Receiving a Subsequent Natural Language Utterance [6B] and [22B] .....                   | 245 |
| 4. | Interpreting Using Long-Term and Short-Term Knowledge [6C], [22C], [32B] and [48B] ..... | 246 |
| I. | Dependent Claims 7, 23, 33 and 49 .....                                                  | 248 |
| J. | Dependent Claims 8, 24, 34 and 50 .....                                                  | 252 |
| K. | Dependent Claims 9, 17, 35 and 43 .....                                                  | 254 |
| L. | Dependent Claims 10, 25, 36 and 51 .....                                                 | 256 |
| M. | Dependent Claims 11, 26, 37 and 52 .....                                                 | 259 |
| N. | Dependent Claims 12 and 38 .....                                                         | 261 |
| O. | Dependent Claims 13, 15, 39 and 41 .....                                                 | 262 |
| P. | Dependent Claims 16 and 42 .....                                                         | 265 |
| 1. | Preamble [16P] / [42P].....                                                              | 265 |
| 2. | Mapping a Stream of Phonemes [16A] / [42A].....                                          | 266 |

**TABLE OF CONTENTS**  
(continued)

**Page vi**

|     |                                                      |     |
|-----|------------------------------------------------------|-----|
| 3.  | Generating a Preliminary Response [16B] / [42B]..... | 267 |
| IX. | Conclusion .....                                     | 268 |

I, Stuart Lipoff, declare as follows:

1. I have been retained by Goodwin Procter LLP on behalf of Petitioners Samsung Electronics Co. Ltd. and Samsung Electronics America, Inc. (“Petitioners”) to provide this Declaration concerning technical subject matter relevant to the petition for *Inter Partes* Review (“Petition”) of U.S. Patent 7,818,176 (“the ’176 patent”; EX-1001).

2. I am over 18 years of age. I have personal knowledge of the facts stated in this Declaration and could testify competently to them if asked to do so.

3. I have been asked to provide my technical opinions regarding how a person of ordinary skill in the art would have understood the claims of the ’176 patent at the time of the alleged invention, which I have been asked to assume is the 2007 timeframe. For purposes of whether the teachings of the prior art render the claims of the ’176 patent obvious, I have been asked to assume the date of February 6, 2007, which I understand to be the filing date of the application that led to the ’176 patent. I have also been asked to provide my technical opinions on how concepts in the ’176 patent specification relate to claim limitations of the ’176 patent.

4. In reaching the opinions provided herein, I have considered the ’176 patent, its prosecution history, and the references cited in my Declaration and the Appendix. I have also drawn on my own education, training, research, knowledge, and personal and professional experience.

5. In general, I have been asked to cite to the specification of a patent or patent publication using the following formats: (Patent, Col:Line Number(s)) or (Patent, Paragraph Number(s)). For example, the citation ('176 patent, 1:1-10) points to the '176 patent specification at column 1, lines 1-10. Also, for convenience, I have been asked to use italics to denote limitations from the challenged claims. Unless otherwise noted, all emphasis is added.

6. I am being compensated for my time at my standard consulting rate. I am also being reimbursed for expenses that I incur during the course of this work. My compensation is not contingent upon the results of my study and analysis, the substance of my opinions, or the outcome of any proceeding involving the '176 patent. I have no financial interest in the outcome of this matter or in any litigation involving the '176 patent.

7. I understand that another party, Amazon.com, Inc., Amazon.com LLC, Amazon Web Services, Inc., A2Z Development Center, Inc. d/b/a Lab126, Rawles LLC, AMZN Mobile LLC, AMZN Mobile 2 LLC, Amazon.com Services, Inc. f/k/a Amazon Fulfillment Services, Inc., and Amazon.com Services LLC (formerly Amazon Digital Services LLC) (collectively “Amazon”), previously challenged the '176 patent in *inter partes* review proceedings at *Amazon.com, Inc. et al. v. VB Assets, LLC*, which was denied institution by the Patent Trial and Appeal Board. No. IPR2020-01390 (P.T.A.B. March 11, 2021) (“Amazon-VBA-IPR”). I refer in my

declaration below to some of the positions and issues raised in that proceeding, including as it pertains to claim construction of certain claim terms in the '176 patent.

## **I. Qualifications**

8. I believe I am well qualified to render useful opinions on this matter. I will briefly summarize my knowledge, training, and experience here. A more detailed summary of my background, education, experience, and publications is set forth in my curriculum vitae (CV), which is submitted as EX-1004.

9. I am currently president of IP Action Partners Inc., a consulting practice that serves the telecommunications, information technology, media, electronics, and e-business industries.

10. I earned a Bachelor of Science degree in Electrical Engineering in 1968 and a second Bachelor of Science degree in Engineering Physics in 1969, both from Lehigh University. I earned a Master of Science degree in Electrical Engineering from Northeastern University in 1974, and then a Master of Business Administration degree from Suffolk University in 1983.

11. I hold a Federal Communications Commission ("FCC") General Radiotelephone Operator License. I also hold a Certificate in Data Processing from the Institute for the Certification of Computing Professionals ("ICCP"), which is supported by the Association for Computing Machinery ("ACM").

12. I am also a registered professional engineer (PE) in the Commonwealth of Massachusetts and in the State of Nevada.

13. I am a fellow of the Institute of Electrical and Electronics Engineers (“IEEE”) Consumer Electronics, Communications, Computer, Circuits, and Vehicular Technology Groups. I have been a member of the IEEE Consumer Electronics Society National Board of Governors (formerly known as the Administrative Committee) since 1981, and I was Boston Chapter Chairman of the IEEE Vehicular Technology Society from 1974 to 1976. I served as the 1996-1997 President of the IEEE Consumer Electronics Society, and from 1999 to 2018 I served as Chairman of the Society’s Technical Activities and Standards Committee and as Vice President of Publications for the Society. From 2018 to 2023 I served as Vice President of Standards and Industry Activities for the Society and currently serve on the Board of Governors as the Historian for the Society. I have also served as an Mazura Ibuka Customer Technology Award committee member for the IEEE’s Award in the field of consumer electronics.

14. I have prepared and presented numerous papers at the IEEE and at other professional meetings. For example, in fall 2000, I served as general program chair for IEEE’s Vehicular Technology Conference on advanced wireless communication technology. I have organized sessions at The International Conference on Consumer Electronics, and I was the 1984 program chairman. I conducted an eight-week IEEE-

sponsored short course on Fiber Optics System Design. I received IEEE's Centennial Medal in 1984, and I received IEEE's Millennium Medal in 2000.

15. As Vice President and Standards Group Chairman for the Association of Computer Users ("ACU") from 1980 to 1983, I served as the ACU representative to the ANSI X3 Standards Group. From 1976 to 1978, I served as Chairman of the task group on user rule compliance for the FCC's Citizens Advisory Committee on Citizen's Band radio.

16. Over the last 25 years, I have been a member of the Society of Cable Television Engineers, the Association for Computing Machinery, and the Society of Motion Picture and Television Engineers. From 2001 to 2004, I served as a member of the USA advisory board to the National Science Museum of Israel. In 1998, I presented a short course on international product development strategies as a faculty member for Technion Institute of Management in Israel. From 2001 to 2003, I served as a member of the board or directors of the Massachusetts Future Problem Solving Program.

17. I am a named inventor on seven United States patents and have several publications on data communications in publications, including Electronics Design, Microwaves, EDN, the Proceedings of the Frequency Control Symposium, Optical Spectra, and IEEE publications.

18. During my over fifty-year professional career, I have been heavily engaged in the study, analysis, evaluation, design, and implementation of products and technology associated with consumer electronics and electronic appliances. A particular focus of my professional activities has been improving the man-machine interface including voice, speech, and speaker recognition for man-machine interactions. I also have extensive experience in foundation technologies and the applications supporting speech signal processing.

19. For approximately three years, from 1969 to 1972, I served as Project Engineer for Motorola's Communications Division, where I had project design responsibilities for paging and wireless communication products. Projects that I worked on while employed at Motorola included work on paging systems that included digital voice storage, voice compression, and voice synthesis. I also worked on projects that interfaced wireless data communications terminals to public safety computer systems for mobile data retrieval and data entry.

20. For approximately four years, from 1972 to 1976, I served as Section Manager for Bell & Howell Communications Company, where I also had project design responsibilities for paging and wireless communication products. I supported projects developing covert audio intelligence systems that recognized speech and activated digital voice compression recording systems. I also led projects for voice-based radio paging systems that recorded speech input, processed the speech to

remove silence, processed the speech to digitally compress the speech, and store and forward the speech upon demand from dual tone multi-frequency or computer keyboard retrieval from the servers.

21. For 25 years from 1976 to 2001, I worked for Arthur D. Little, Inc. (“ADL”), where I became the Vice President and Director of Communications, Information Technology, and Electronics (CIE) and served in that role for 10 years, from 1991 to 2001. At ADL, I was responsible for the firm’s global CIE practice in laboratory-based contract engineering, product development, and technology-based consulting. I was also involved in multiple pioneering efforts to identify and explore customer-to-business and business-to-business electronic commerce and transactions information processing opportunities, or “e-commerce”. These projects involved technology assessment and analysis as well as developing architectures and systems to support multiple applications. The technologies typically involved an information retrieval component.

22. While at ADL, I worked on several projects involving the combination of voice interfaces (including speech recognition and voice audio output) and information retrieval. Over the course of three years in the early-1990s, I worked on a project for Bolt Beranek and Newman (BB&N), where I evaluated and benchmarked technology for a voice input/output application that allowed end users (*e.g.*, travel agents) to use speech inputs to interact with airline reservation databases

to retrieve information about travel reservation options, where the results were returned to the user in an audible message. This system included a natural language front-end speech-interface module with speech recognition that used pre- defined recognition grammars to convert the end user's speech into structured commands supported by an airline reservation system. As another example, over the course of three years in the mid-1990s, I also worked on a project for Texas Instruments that applied a speech-recognition interface for a variety of applications that retrieved information from database servers.

23. Other projects that I worked on at various points in my twenty five years at ADL and afterwards involving speech recognition technologies including systems for Sprint, Rockwell Semiconductor, and the United States Postal Service ("USPS"), detailed below.

24. Over the course of three years in the early 1990s, I worked on a voice-interface project developing spoken digit telephone number recognition and voiceprint matching for Sprint's long distance alternative access telephone services.

25. Over the course of a year in the late 1980s, I worked on a voice interface project evaluating the processing power needed to perform various voice recognition applications by Rockwell Semiconductor's signal processing technology.

26. Over the course of 15 years starting in the early 1980s, I worked on a project for the USPS, where we developed a real-time automated postal teller system

that served as an interface between end-users and the USPS's information systems. This system included voice prompts for the vision impaired.

## **II. Understanding of Relevant Legal Principles**

27. I am not a lawyer, and I will not provide any legal opinions. Although I am not a lawyer, I have been advised certain legal standards are to be applied by technical experts in forming opinions regarding the meaning and validity of patent claims.

28. I understand that a patent claim is invalid if it is anticipated or obvious in view of the prior art, and that a claim can be unpatentable even if all of the requirements of the claim cannot be found in a single prior-art reference. I further understand that invalidity of a claim requires that the claim be anticipated or obvious from the perspective of a person of ordinary skill in the art at the time the invention was made.

29. I have been informed that a patent claim is invalid if it would have been obvious to a person of ordinary skill in the art. In analyzing the obviousness of a claim, I understand the following factors may be taken into account: (1) the scope and content of the prior art; (2) the differences between the prior art and the claims; (3) the level of ordinary skill in the art; and (4) any so called "secondary considerations" of non-obviousness, if they are present. I am not aware of any evidence of secondary considerations of non-obviousness relevant to the '176 patent.

I reserve the right to supplement this Declaration if Patent Owner (“PO”) introduces evidence of secondary considerations of non-obviousness.

30. I understand that to prove that prior art or a combination of prior art renders a patent obvious, it is necessary to:

- (1) identify the particular references that, singly or in combination, make the patent obvious;
- (2) specifically identify which elements of the patent claim appear in each of the asserted references; and
- (3) explain why a person of ordinary skill in the art would have combined the references, and how they would have done so, to create the inventions claimed in the patent. I further understand that exemplary rationales that may support a conclusion of obviousness include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- use of known technique(s) to improve similar devices (methods or products) in the same way;
- applying a known technique to a known device (method or product) ready for improvement to yield predictable results;
- “obvious to try” – choosing from a finite number of identified, predictable solutions with a reasonable expectation of success; known work in one field of endeavor may prompt variations of the work for use in either the same field or a different field based on

design incentives or other market forces if the variations are predictable to a person of ordinary skill in the art; and

- some teaching, suggestion, or motivation in the prior art that would have led a person of ordinary skill in the art to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

31. I have been informed that, in considering obviousness, hindsight reasoning derived from the patent-at-issue may not be used.

### **III. Technical Background**

#### **A. Language and Dialog**

32. Spoken language “is very different to written language” because “people typically do not follow rigid syntactic and morphological constraints in their utterances.” (Bernsen; EX-1019, 49.) The “lack of written language formality in spontaneous spoken language makes linguistic analysis by machine both more difficult than, and different to, analysis of written language.” (Bernsen, 49.)

33. As Huang explains “[i]t is natural to assume that the *sentence* is a clear and simple chunking unit for dialog, by analogy with written communication.” (Huang; EX-1020, 860 (emphasis in original).) But, because “sentences are artificially delimited in written text, researchers in dialog communication usually speak of the *utterance* as the basic unit.” (Huang, 860 (emphasis in original).)

34. In the context of conversational speech, one or more utterances can comprise a “turn.” A turn in conversation refers to the time one speaker holds the

floor, often including multiple utterances—the smaller units of speech. Turns structure the flow of dialogue, beyond just the speech segments themselves. Turns are viewed as “building blocks for constructing a common task-oriented understanding among participants.” (Huang, 861.)

35. Dialog is best understood not as a series of isolated turns but as structured discourse in which utterances are interdependent and contextually grounded. A dialog “consists of at least two turns, one by each speaker”, and a coherent dialog “will exhibit discourse phenomena” whose interpretation “depends on the dialogue context.” (McTear; EX-1021, 46.)

36. Two examples of how dialog presents these issues are anaphora and ellipsis. Anaphora is a linguistic phenomenon where a word or phrase refers back to another word or phrase mentioned earlier in the discourse. In dialog systems, anaphora resolution is essential for understanding references like pronouns. Ellipsis occurs when part of a sentence is omitted because it is understood from context. In dialogue, users often leave out information that has already been established, and the system must infer the missing parts. (Kehler; EX-1022, 62, 64.)

37. A key structuring device in dialogue is the adjacency pair—predictable turn sequences like question–answer or greeting–greeting—that capture mutual expectations and conversational obligations. (McTear, 65). These pairs are not merely formalities but reflect deeper discourse coherence, and they often aggregate

into larger goal-driven patterns known as dialog games. (McTear, 65-68.) These games model the intentional structure of conversation and accommodate embedded exchanges, such as clarifications or corrections, within a larger purpose. (McTear, 65-68.)

38. Common terminology exists in the field of interactive/spoken dialog systems. For example, a “*lexicon* is a list of words, a *vocabulary*, annotated with syntactic (including morphological) and semantic features.” (Bernsen, 49 (emphasis in original).) “*Grammars* describe how words may be combined into phrases and sentences.” (Bernsen, 50 (emphasis in original).) “*Semantics* are abstract representations of the meanings of words, phrases and sentences.” (Bernsen, 50 (emphasis in original).)

## **B. Interactive/Spoken Dialog Systems**

39. At a high level, the task of an interactive dialog system “is to enable and support the spoken interaction between the human user and the service offered by the application system.” (Möller; EX-1023, 19.) This high-level tasks leads to a number of sub-tasks performed by the system, including:

- coherence of the user input has to be verified, taking into account linguistic and task- or domain-related knowledge;
- communicative and task goals have to be negotiated with the user, and problems occurring during the interaction have to be resolved;
- references like anaphora or ellipses in the user’s utterances have to be resolved;

- inferences which are reasonable in the communicative and task context have to be drawn, and the most probable user reaction has to be predicted.

(Möller, 19.)

40. Möller generally presents four types of systems that can be differentiated “[d]epending on the complexity of the dialogic interaction”:

- **Command systems:** “They are characterized by a direct and deterministic interaction. To each stimulus from one agent corresponds a unique response from the other. The response is independent of the state or context of each agent. This type of interaction is normally not considered as a dialogue, and is called a ‘tool metaphor’. Example: Pressing a key on the keyboard results in a character appearing on the screen.” (Möller, 19.)
- **Menu dialog systems:** “To this class belong simple question-answer user interfaces, where dialogue and task models are merged. The interaction is mainly system-directed, permitting only very little user initiative (e.g. barge-in). In contrast to command systems, several exchanges may be necessary in order to provoke one action of the application system. On the other hand, one user input can provoke different responses, depending on the internal state of the system, e.g. the current level in the menu structure. Example: So-called ‘Interactive Voice Response’ (IVR) systems which enable an interaction via

Dual Tone Multiple Frequency (DTMF) or keyword recognition.” (Möller, 19.)

- **Spoken Dialog System (SDS):** “This narrow class of systems disposes of distinct and independent models for task, user, system, and dialogue. Context information is taken into account using a particular knowledge base or dialogue history. Multiple types of references can be processed. An SDS may be capable of reasoning, of error or incoherence detection, internal correction, anticipation, and prediction.” (Möller, 19-20.)
- **Multimodal dialogue systems including speech:** “Systems of this class show the same characteristics as SDSs do, but in addition they are able to process and synchronize different modality information. Appropriate modalities for the individual interactions have to be selected, both at the system input and output side.” (Möller, 20.)

41. An interactive/spoken dialog system has a “modality” which “is simply a form (or mode) of representing information as output from, or input to, a computer system (Hovy and Arens, 1990).” (Bernsen, 77.) Some systems only support a single modality (e.g., speech) whereas others support multiple and are referred to as multimodal systems.

42. Spoken dialog systems, including multimodal dialog systems, are “an interface between a human user and an application system which uses speech as the

interaction modality (Fraser, 1997).” (Möller, 18.) As such, the interface must be able to process information “coming from and going to the user through the speech-technology-based interface (voice user interface, VUI)” and information “coming from and going to the application system through a specialized (e.g., SQL-based) interface.” (Möller, 18.) The system “must achieve a number of actions in order to be able to give a response, and the response will depend on the internal state of the system, or on the context of the interaction.” (Möller, 18.)

43. Most early spoken dialogue systems were designed for and operated over the telephone network, largely due to the widespread availability and standardized nature of telephony infrastructure. These systems typically employed a phone server interface to manage user access and interactions via voice calls, making them suitable for tasks like automated customer service and information retrieval (Möller, 20.)

44. A spoken dialog system “consists of six major components which are accessed by the user” via an access device. (*See, e.g.,* Möller, 20.) These are:

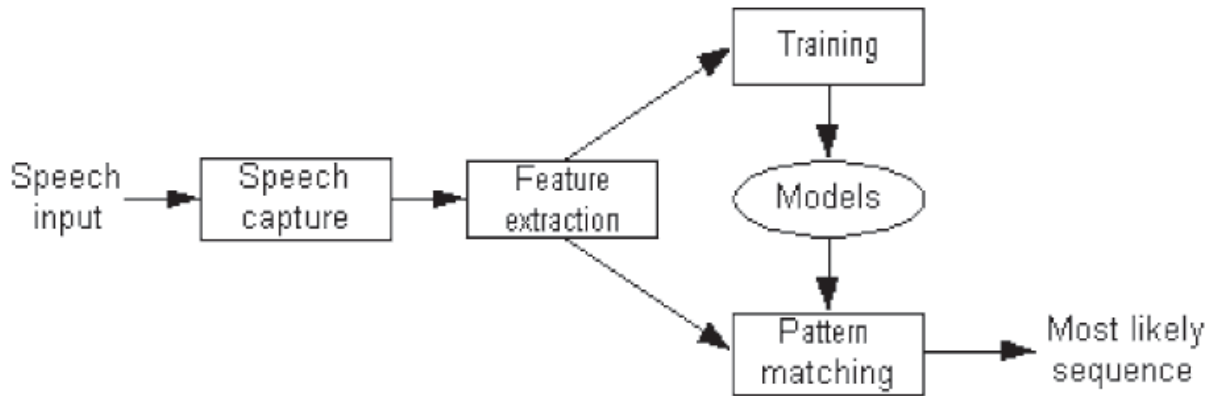
- **Phone Server:** Provides access to the user over the telephone network.
- **Speech Recognizer:** Converts the spoken input into a word string or word hypothesis graph.
- **Semantic Analyzer:** Interprets the recognized word string semantically, producing a semantic frame.

- **Dialogue Manager:** Interprets the semantic frame in context, manages dialogue history, and controls dialogue flow.
- **Response Generator:** creates natural language replies based on the dialogue manager's output.
- **Speech Synthesizer:** Converts the response into speech, and transmits it back to the user.

45. In the next section I describe what are the key components of any speech dialogue system: the speech recognizer, the semantic analyzer, the dialogue manager, and the response generator and speech synthesizer.

### 1. Speech Recognition Component

46. The speech recognition component is commonly referred to as an automatic speech recognizer (ASR). The speech recognition process is illustrated by McTear's Figure 4.2 (below). Speech recognition begins when a user "speaks into a microphone or telephone handset." (McTear, 83.) The analog acoustic signal "is captured by the microphone and converted to an electric current, which is then passed to an analogue-to-digital converter within the sound card that translates the current into a stream of bits that represent the sound." (McTear, 83.) This part of the process, known as signal processing, "extract[s] the relevant set of features from the acoustic signal." (McTear, 83.)



**McTear, Figure 4.2**

47. The extracted features are then “classified as sequences of phonemes using an acoustic model, and these sequences of phonemes are then combined into words that represent the system’s estimate of what the user said.” (McTear, 83.) The process of phoneme and word identification involves an acoustic model “that shows how each word consists of a sequence of phonemes and how each phoneme relates to the values of the features extracted from the acoustic signal” and a “language model that specifies permissible sequences of words.” (McTear, 84.)

48. “The acoustic model captures variability in pronunciation using probabilities.” (McTear, 84.) Words may be pronounced in different ways by different individuals. (*See, e.g.*, McTear, 84.) For example, the word “tomato” may be pronounced “T AH M EY T OW” by American speakers or “T AH M AA T OW” by British speakers. (McTear, 84.) Hidden Markov Models (“HMMs”) are used in

speech recognition systems “to represent [] variable patterns of speech.” (McTear, 85.)

## 2. Language Understanding

49. The language understanding component “analyse[s] the output of the speech recognition component” and “assign[s] a meaning representation that can be used by the dialogue manager.” (McTear, 91.) Traditionally, language understanding involved “syntactic analysis, to determine the constituent structure of the recognised string (i.e., how the words group together), and semantic analysis, to determine the meanings of the constituents.” (McTear, 91.) Language understanding is problematic because of ambiguity in natural language and ill-formed input. (McTear, 91.)

50. McTear lists the following ways in which natural language can be ambiguous:

- **Lexical ambiguity:** “A word may belong to more than one part of speech, for example, ‘book’ can be a noun or a verb.” (McTear, 91.) This type of ambiguity “can usually be resolved within the context of the other words in the sentence, as in the sentence ‘book a flight to London’.” (McTear, 91.)
- **Sense ambiguity:** “A word can have different meanings, for example, ‘bank’ can be a financial institution or the side of a river.” (McTear, 91.)
- **Structural ambiguity:** “The relationship between the phrases in the sentence is ambiguous, for example, ‘a flight to London arriving at 9.’ On a purely

syntactic analysis, there are two possible readings, one in which the flight arrives at 9, and the other in which London arrives at 9.” (McTear, 91.)

51. In typical systems, the semantic analyzer uses predefined grammars, slot-filling methods, or statistical models to extract relevant information from user utterances. These might include details such as dates, locations, or actions, which are organized into semantic frames or attribute-value structures. For instance, in a travel booking system, an utterance like “I want to fly to Berlin next Monday” would be converted into a frame containing values for destination and travel date. (*See Möller, 29-30.*)

52. The language understanding component will often rely on a component known as a parser to assist in this process. The parser is a tool or algorithm that analyzes the grammatical structure of the recognized speech input. Its role is to break down the sentence into its syntactic components (like subjects, verbs, objects, etc.) and determine how those components relate to each other. This helps the system understand what the user means, not just what words were spoken. For example, if a user says, “Book a flight to Paris next Friday,” the parser helps identify that:

“Book” is the action (verb),

“a flight” is the object,

“to Paris” indicates the destination,

“next Friday” refers to the date.

This structure can then be processed by the dialogue manager to determine the next system action. (*See Möller, 26, 29-30.*)

### **3. Dialog Management Component**

53. In a spoken dialog system, the dialog manager is the “central component.” (McTear, 107.) The Dialog Manager “accepts spoken input from the user, produces messages to be communicated to the user, interacts with external knowledge sources, and generally controls the dialogue flow.” (McTear, 107.) Put differently, according to Möller, “It is the task of the dialogue manager to guarantee the smooth course of the dialogue, so that it is coherent with the task, the domain, the history of the interaction, with general knowledge of the ‘world’ and of conversational competence, and with the user.” (Möller, 27.)

54. Dialog managers access both short-term and long-term knowledge to operate effectively. Short-term knowledge, sometimes referred to as short-term memory, often is used to refer to information particularly pertinent to a particular session whereas long-term knowledge, sometimes referred to as long-term memory, often is used to refer to information that would be relevant across multiple user sessions. Because of this characteristic of these two types of knowledge, it is common for short-term knowledge to be non-persistent and long-term knowledge to be persistent. Dialog managers, like all computational systems, have incorporated

the concepts of short-term and long-term knowledge long before the priority date of the '176 patent.

#### **4. Response Generator and Speech Synthesizer**

55. The response generator and speech synthesizer in a spoken dialogue system is responsible for converting internal representations—typically derived from the dialogue manager—into linguistically and pragmatically appropriate output. “They are described together, because the strict separation into a component which generates a textual version of the output for the user (response generation) and another one which generates an acoustic signal from the text (speech synthesizer) is not always appropriate.” (Möller, 33.)

56. As Möller explains, “[r]esponse generation involves decisions about what information should be given to the user, how this information should be structured, and about the form of the message (words, syntax).” (Möller, 33.) The system must decide not just what to say, but how to say it—balancing informativeness, clarity, and conversational appropriateness. This process may be implemented via formal grammars or simpler “template” approaches, in which sentences are generated by filling predefined structures with contextually relevant content.

57. Möller further notes that at each dialogue act, “the response generator builds a template sentence ... filling gaps from the content of the current semantic

frame, the dialogue history, and the result of the database query.” (Möller, 33.) Higher-level generation rules govern how much information can be included in a single utterance and how it should be structured, particularly when there is too much content to communicate all at once. Importantly, response generation should be informed by the user model, adapting language and content based on the user’s background knowledge and experience. As Möller puts it, this module must “respect the user model, e.g. with respect to his/her expected domain knowledge and experience.” (Möller, 33.) Ultimately, the response generator ensures that the system communicates in a way that is coherent, concise, and aligned with user expectations.

### **C. Intent-Based Advertising**

58. Intent-based advertising refers to the strategic targeting of individuals who are actively seeking information—often in relation to a specific product or purchase decision. This method of advertising hinges on the ability to discern commercial intent from user behaviors, particularly search activity. (Dai; EX-1024, 1.) As Dai describes it, understanding whether a person is “about to commit or in the middle of a commercial activity” can significantly enhance the relevance and profitability of content delivery across the web. (Dai, 1.)

59. The concept is not new: traditional mediums like newspaper classifieds and trade directories have long allowed advertisers to place messages directly in front of consumers demonstrating clear intent to engage in commercial activity. The

evolution continued with the widespread adoption of telephone directories, most notably the Yellow Pages. The Yellow Pages became an iconic example of intent-based advertising in the pre-digital era, allowing businesses to list themselves in categories that matched consumer searches precisely. People who needed a plumber, restaurant, or mechanic would actively consult these directories, creating a natural alignment between consumer intent and advertised services.

60. The rise of digital technology led to services like 1-800-FREE-411 in the early 2000s, an innovative phone-based directory assistance service that provided business listings free to callers. Funded by brief audio advertisements, this service tapped directly into consumer intent—callers explicitly seeking information about businesses or services received relevant advertisements. By embedding advertising into directory assistance calls, 1-800-FREE-411 effectively anticipated and served immediate consumer needs. Similar services included GOOG-411, a free voice-activated search tool provided by Google, and Tellme Networks, which also offered voice-driven business search with ad-supported content. These services bridged the gap between analog information-seeking and digital convenience, all while monetizing the moment of user intent.

61. While the idea is not new, technology provides tools to improve the precision with which digital platforms can now detect and respond to these signals of interest. Indeed, as of the earliest priority date of the '176 family patents,

researches were actively studying how to improve such tools. For example, Dai describes numerous frameworks for determining commercial intent, such as:

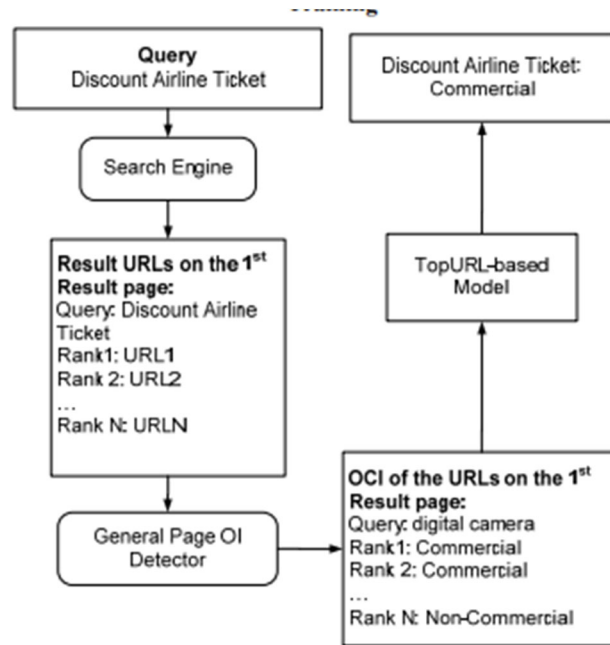


**Figure 5: Detect OCI based on Search Result Landing Pages - Prediction**

**Dai, Figure 5**

#### **D. Web-Based Advertising Tools**

62. Some of the fastest growth of web-based advertising took place between the years of 1995 and 2005, where total online ad revenue increased from approximately \$55 million to \$12.5 billion, an increase of over 200x. (IAB; EX-1025, 11.) As discussed further below, the period between 1995 and 2005 corresponded with the development of new technologies that took advantage of the unique capacity of web-based platforms to provide consumers with contextually

appropriate advertisements that were suited for the particular consumer. (Plummer; EX-1026, 7-29.)

63. This technology addressed a long-standing practical issue in advertising. As one investigator observed in 2004, “[e]nsuring that media is bought effectively and not directed towards the ‘wrong people’ has always been a challenge for marketers.” (Iyer; EX-1027, 1.) Thus, advertisers’ “objectives are often to target advertising to specific consumer groups.” (Iyer, 1.)

### **1. Early Web-Advertising Systems**

64. Around 1995, early websites and search engines, such as Yahoo! and AltaVista, offered basic forms of advertising that matched ads to the general topic of a page. For instance, a user browsing a page about travel might see banner ads for airline deals or hotel bookings. (*See* Plummer, 9.) These ads were relatively unsophisticated, relying primarily on static keywords and general page categories rather than user behavior or preferences. They were also often manually placed and lacked precision.

65. A key development in these early systems came with ad platforms like DoubleClick, which started in 1996 (and was bought by Google in 2007). These platforms allowed advertisers to not only deploy ads across many websites but to also systematically track the effectiveness of ads. As Google described in the press release announcing the acquisition: DoubleClick’s ad servers placed ads in the “right

location at the right time” and reported real-time performance, which allowed advertisers to “change the content, and timing of ads almost on the fly.” This made it possible to “maximize the return on investment for a given ad or campaign.” (Google Blog; EX-1028, 4.)

66. Another development from this period was the development and implementation of tools that allowed search-based targeting, sometimes called “sponsored search.”

67. Sponsored search is a form of online advertising in which advertisers pay to have their messages displayed alongside search engine results, targeting users based on the keywords they enter. These features enabled search engines to offer their core services for free while generating substantial revenue through paid listings. As Fain and Pedersen explain, sponsored search integrates “relevance-targeted text advertisements” directly into the search experience, and by 2005, search engines were delivering over 13 billion sponsored results per week. (Fain; EX-1029, 1.)

68. The model is built on several key components: advertiser-submitted content (titles, descriptions, and URLs) tagged with relevant keywords; a bidding system where advertisers set prices they are willing to pay per click; and an auction-based ranking mechanism that determines ad placement based on both bid value and, in many systems, performance metrics such as click-through rate (CTR). (Fain, 2-3.)

69. GoTo.com, later known as Overture Services, was the first to successfully deploy this model in 1998. It allowed advertisers to bid for placement on specific keywords, and search engines would display these listings in rank order based on the bid amount. (Fain, 1-2.) Relevance was ensured through a combination of editorial review and, increasingly, automated algorithms.

70. The model quickly gained traction. Google's adaptation of sponsored search in 2002 introduced a significant innovation: ranking ads not solely by bid amount but by the product of bid and CTR, thereby rewarding ads that were not only lucrative but also engaging to users. (Fain, 3.) This reflected one of the earliest ways in which advertising engaged in a feedback loop based on which ads drove customer clicks. Yahoo!, having acquired Overture in 2003, continued to refine the system by introducing the "Click Index," a measure comparing a listing's observed CTR to expected norms. (Fain, 3.)

71. Beyond basic cost-per-click (CPC) pricing, the industry expanded to include cost-per-action (CPA) and cost-per-mille (cost per thousand impressions or CPM) models. These enabled advertisers to align their expenditures more closely with measurable outcomes, such as purchases or signups, and allowed publishers to better monetize their platforms. (Fain, 2.) With the rise of behavioral targeting, time-of-day bid adjustments, and sophisticated analytics, sponsored search evolved into a dynamic, data-driven marketing platform.

72. One of the most famous instantiations of this approach was Google Adwords, which launched in 2000. (Google Press Release; EX-1030, 1.) AdWords enabled advertisers to “purchase individualized and affordable keyword advertising” that appeared “instantly on the google.com search results page.” (Google Press Release, 1.) With this system, ads were “highly targeted,” triggered only when a user entered a search query that matched an advertiser’s “carefully selected keywords.” (Google Press Release, 1-2.) As Google emphasized, this approach “ensures advertisers will be delivered a prequalified visitor and will not pay for ‘off target’ impressions.” (Google Press Release, 2.)

73. The system also offered detailed “web-based reporting tools,” including real-time “click-through rates and number of impressions delivered,” and “visitor interest bar[s]” to provide immediate feedback on ad effectiveness. (Google Press Release, 2.) Ads appeared as “quick-loading AdWords text ads ... to the right of the Google search results.” (Google Press Release, 2.)

## **2. User Modelling Approaches**

74. By 2007, basic data analysis tools and growing web tracking methods had been developed that allowed websites to gather information from which advertisers could build a model of who was viewing the webpage. Using these models, advertisers could then substantially improve their ability to show ads that were likely to get a response. As Plummer explains, such targeting methods have

been shown to significantly improve key performance indicators such as click-through and conversion rates, offering a clear incentive for advertisers to invest in increasingly granular audience segmentation. (Plummer, 13.)

75. The inventors of the '176 Family of patents describe three “models” that can similarly inform the selection of an advertisement: The “personalized cognitive model,” the “generalized cognitive model,” and the “environmental model.” These models track onto three broad categories of modelling used in advertising, including web-based advertising.

76. The '176 patent claims an environmental model “derived from environmental conditions or surroundings associated with the specific user.” ('176 patent, claims 2, 18, 28, 44.)

77. The claimed “environmental model” closely tracks a range of advertising strategies that deliver messages based on attributes of a user’s immediate physical or temporal context. Two primary examples are geotargeting—the use of location data to determine where a user is at the moment an ad is served—and daypart targeting, which tailors ad delivery according to the time of day. (Plummer, 17-24.)

78. The '176 patent also claims a personalized cognitive model “derived from an interaction pattern for a specific user.” ('176 patent, claims 2, 18, 28, 44.)

79. This claimed model closely tracks the idea of behavioral targeting, which refers to the practice of delivering advertising based on inferences drawn from a specific user's prior online activity. (Plummer, 11-16.) This activity may include websites visited, search queries submitted, purchases made, or other interactions that form a behavioral profile. The goal is to increase ad relevance by aligning content with the user's demonstrated interests, rather than their current context or demographic characteristics. (*See* Plummer, 11-12; *see also* Duane Morris: EX-1031.)

80. This method emerged as a key innovation in digital advertising in the early 2000s, as improvements in browser tracking, data storage, and network-level advertising allowed for more sophisticated observation of user behavior. According to Duane Morris LLP, behavioral targeting became feasible through the collection of clickstream data, which enabled advertisers to infer a user's interests without requiring that the user explicitly provide personal data. (Duane Morris, 2). This approach allowed networks to serve ads that were customized based on individual behavioral signals, rather than generalized audience traits.

81. Plummer describes behavioral targeting as one of the most powerful tools available to online advertisers, noting that networks can deduce user profiles based on patterns of activity across unrelated websites. For instance, a user who visits several indie music blogs, parenting forums, and auto comparison sites in a

short time span may be categorized as having multiple distinct interests—each of which can trigger targeted ad delivery. (Plummer, 12 (citing industry practice circa early 2000s).) Plummer also cites empirical findings indicating that behaviorally targeted ads can achieve significantly higher engagement than standard or contextual ads. One cited study found that such ads had “108% higher click-through rates than in-context advertising” and a 19% increase in post-click conversion rates, suggesting not just improved attention but also increased action. (Plummer, 13 (citing Tacoda and Revenue Science data, ca. 2004–2006).)

82. Plummer further situates behavioral targeting within a broader strategic framework. It explains that behavioral methods can be used both for acquisition (e.g., targeting users in the early stages of purchase consideration) and for retention (e.g., re-targeting users who abandoned a shopping cart). Plummer emphasizes that the strength of behavioral targeting lies in its ability to match ad messaging to users’ demonstrated, rather than assumed, interests—effectively bypassing the limitations of self-reported data or static segment definitions.

83. The ’176 patent also claims a “generalized cognitive model” that is “derived from an interaction pattern for a plurality of users.” (’176 patent, claims 2, 18, 28, 44.)

84. This claimed model closely tracks the idea of Demographic Targeting, or the delivery of advertising based on consumer characteristics such as age, gender,

household income, education level, or occupation. (Plummer, 8-9.) These attributes are typically inferred or obtained through user registration data, third-party data aggregators, or predictive modeling. The central premise is that certain demographic groups exhibit consistent preferences or purchasing behaviors, making them suitable proxies for audience selection.

85. Plummer identifies demographic targeting as one of the earliest and most established forms of audience segmentation. (Plummer, 8.) In the context of online advertising, it represents a digital extension of traditional media planning practices where ad buys were matched to the assumed demographics of television programs, radio shows, or print publications. Online platforms adapted these methods by either collecting self-reported data (e.g., during account creation) or applying statistical models to infer demographic attributes from behavioral or contextual signals. (Plummer, 8-9.)

### **3. Using Analytics to Improve Models**

86. Burns noted in the early 2000s that “[a]ll the search engines track click-through rates (CTRs) and impression data for each of your keywords and creatives.” (Burns; EX-1032, 251.) These metrics serve as a foundational component of modern advertising systems, capturing both the visibility of advertisements (impressions) and the corresponding user responses (clicks). The systematic collection of this data enables advertisers and platforms to evaluate ad performance and optimize delivery

strategies. Beyond campaign reporting, this interaction data provides a valuable input for refining the user models described in the previous section, contributing to ongoing adjustments in how audiences are segmented and targeted.

87. User profiles in digital advertising systems are constructed through the aggregation of interaction data over time. These interactions may include ad clicks, search queries, content views, and responses to promotional offers. Systems use this data to develop statistical models that estimate a user's interests, preferences, or likelihood to engage with specific content. Plummer explains that advertising networks such as Tacoda and Revenue Science observe behavioral patterns across multiple unrelated sites to infer topic-level interests. These observations allow the system to assign users to affinity segments—clusters defined by similar patterns of interaction—without requiring explicit input from the user. (Plummer, 12-13 (citing Tacoda and Revenue Science).) The result is a dynamic profile that evolves with continued interaction and is used to inform future ad selection and delivery.

88. This profiling process forms the basis of behavioral targeting systems, which rely on interaction data to infer user preferences and predict future behavior. As described in the previous section, behavioral targeting involves constructing statistical models that associate observed actions—such as ad clicks or content consumption—with underlying interests. These models are not static; rather, they are continuously updated as new interaction data becomes available. Each new data

point serves to refine the system's estimates, enabling more accurate classification of user interests and improving the relevance of subsequent ad delivery. This iterative updating mechanism is central to maintaining model accuracy over time and adapting to shifts in user behavior. As Plummer describes, the behavioral approach is not just about making ads more relevant, the approach also allows companies to track and analyze customers to "profile[] each by their brand interactions." (Plummer, 53 (citing Anfuso 2005).) The result of this is that the company's "database will only become richer over time as behavioral tracking continues with each successive year." (Plummer, 54.)

#### **IV. '176 Patent**

89. The '176 patent is titled "System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input" and is generally directed to "a system and method for selecting and presenting advertisements based on natural language processing of voice-based inputs." ('176 patent, Abstract.)

##### **A. Overview of the '176 Patent**

90. The '176 patent claims priority to non-provisional U.S. Application No. 11/671,536, filed on February 6, 2007. Although I do not have an opinion about whether the '176 patent is entitled to this priority date, each applied reference was filed or published before February 6, 2007.

91. The '176 patent is titled “System and Method for Selecting and Presenting Advertisements Based on Natural Language Processing of Voice-Based Input” and is generally directed to “[a] system and method for selecting and presenting advertisements based on natural language processing of voice-based inputs.” ('176 patent, Abstract.)

92. Figure 3 of the '176 patent, which I reproduce below, is an embodiment of the disclosed method for selecting and presenting advertisements ('176 patent, 6:66-7:1):

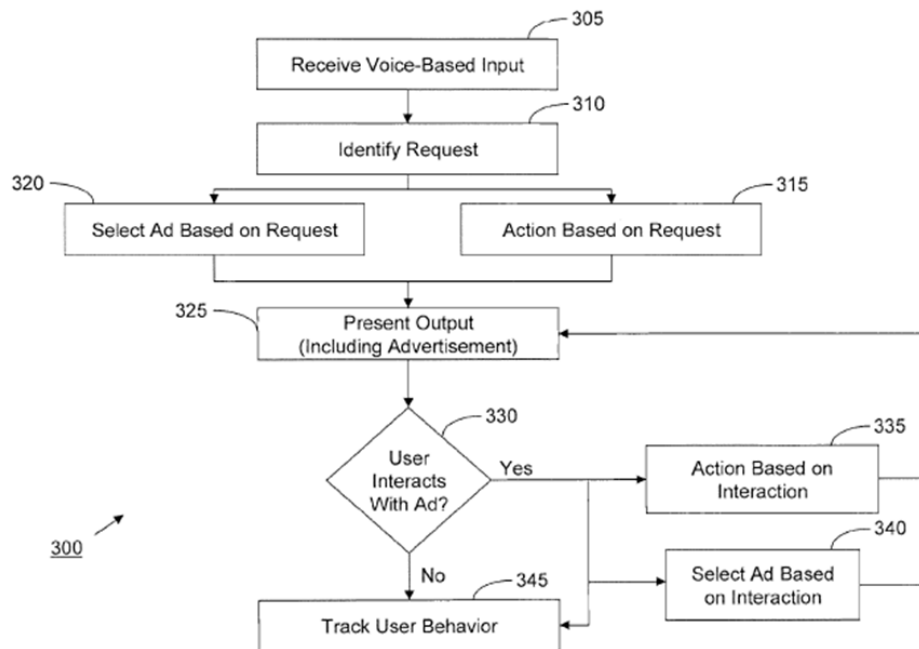


Fig. 3

### **'176 Patent, Figure 3**

93. Figure 3 of the '176 patent teaches “an operation 305, where a voice-based input, including at least a user utterance” is received at a “voice-user interface” ('176 patent, 7:1-4.) The voice-based input may include “multiple requests,” including “requests to retrieve information, perform tasks, explore or gather information, or otherwise interact with a system or device.” ('176 patent, 7:10-14, 7:25-31.)

94. Operation 310 parses the voice-based input in operation 305 by identifying the request in the voice-based input. ('176 patent, 7:10-11.) The request can be interpreted “based upon previous utterances in a current conversation, utterances in previous conversations, context-based information, local and/or global user profiles, or other information.” ('176 patent, 7:31-37.)

95. “Upon identifying the one or more requests,” Figure 3 shows that “[1] action may be taken based on the identified requests in an operation 315, while [2] one or more advertisements may be selected in an operation 320.” ('176 patent, 7:58-61.) Operation 320 (i.e., selecting an ad based on the request) may employ various techniques to select advertisements. ('176 patent, 10:10-13.) Operation 320 may also use various statistical profiles to “define affinities between advertisements, topics, users, etc. (e.g., based on click-through or conversion rates, or other tracking information ...).” ('176 patent, 10:20-23.)

96. Figure 3 of the '176 patent shows that “upon taking action in operation 315 (e.g., to serve the request) and selecting one or more advertisements in operation 320 (e.g., in relation to the request), an output may be presented to the user in operation 325.” ('176 patent, 8:6-10.) The output in operation 325 “may include requested information, an indication of whether a requested task was successfully completed, whether additional information is needed to service the request (e.g., including a prompt for the information), or other information relating to an action based on the request.” ('176 patent, 8:11-17.) “[T]he output may also include advertisements, as selected in operation 320.” ('176 patent, 8:17-18.)

97. “The output in operation 325 may also be presented to the user in various ways.” ('176 patent, 10:30-51.) “[I]n various implementations, the output may include a voice-based or otherwise audible response.” ('176 patent, 10:31-33.) The output may also be “displayed on the display device.” ('176 patent, 10:34-35.)

98. The bottom portion of Figure 3 of the '176 patent shows that “[s]ubsequent interaction between the user and the presented advertisements may be monitored in a decisional operation 330.” ('176 patent, 10:52-54.) Further tracking takes place in operation 345. ('176 patent, 11:11-12.) “For example, operation 345 may track historical data about users, conversations, topics, contexts, or other criteria to associate information with the selected advertisement.” ('176 patent, 11:12-15.)

99. Figure 3 of the '176 patent further shows that “when the user elects to interact with the advertisement, action may be taken based on the interaction in an operation 335.” ('176 patent, 10:54-56.) The user’s interaction with the advertisement may take various forms, including clicking on the advertisement displayed or requesting action from a voice-enabled media device. ('176 patent, 10:52-11:4.) “[A]dditional advertisements may be selected in an operation 340 based on the interaction, using similar techniques as described in connection with operation 320.” ('176 patent, 11:4-10.)

100. There are two other figures in the '176 patent. Figure 1 of the '176 patent, which I reproduce below, shows an “exemplary system 100 for implementing a voice user interface.” ('176 patent, 2:61-63.)

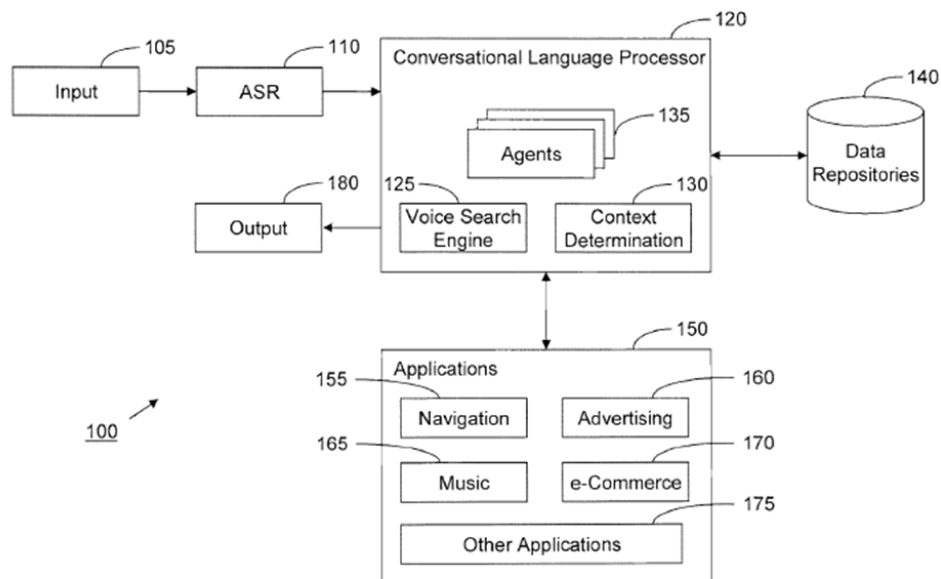


Fig. 1

### **'176 Patent, Figure 1**

101. The '176 patent teaches that System 100 may be used to “control navigation devices, media devices, personal computers, personal digital assistants, or any other device supporting voice-based inputs.” ('176 patent, 2:63-67.)

102. System 100 includes an input device 105 that receives a user input, “including at least a voice-based user utterance.” ('176 patent, 3:28-30.) Input devices may include a microphone or other multi-modal input, such as a touch-screen interface, keypad, or other input. ('176 patent, 3:30-35.)

103. System 100 includes a speech recognition engine 110, which the '176 patent refers to as an “Automatic Speech Recognizer” or “ASR.” The ASR processes the utterance received at the input device 105 and generates one or more “preliminary interpretations” of the utterance using “various techniques,” including “phonetic dictation to recognize a stream of phonemes,” ('176 patent, 3:35-37), or “dictionary or phrase tables,” '176 patent, 3:16-19.

104. The ASR's preliminary interpretations pass to a conversational language processor 120. ('176 patent, 3:52-53.) As shown in Figure 1 of the '176 patent, the conversational language processor 120 includes “a voice search engine 125, a context determination module 130, and one or more agents 135, among other things.” ('176 patent, 3:53-57.)

105. The context determination module 130 “establish[es] a context for the current utterance 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).” (’176 patent, 4:25-34.) The domain agents 135 can model various domains, including navigation, music, a specific user, global users, advertising, and e-commerce. (’176 patent, 4:30-34.)

106. The conversational language processor 120 is also capable of “adaptive misrecognition,” meaning that it may reinterpret a current utterance or one or more previous utterances. (’176 patent, 4:43-45.)

107. The conversational language processor 120 may be “communicatively coupled” to “one or more data repositories 140 and one or more applications 150.” (’176 patent, 4:58-60.) The applications 150 may include a navigation application 155, an advertising application 160, a music application 165, an electronic commerce application 170, and/or other applications 175. (’176 patent, 4:5-9.) The data repositories 140 and the applications 150 “may reside locally (e.g., on a user device), remotely (e.g., on a server), and/or hybrid local/remote processing models may be used (e.g., lightweight applications may be processed locally while computationally intensive applications may be processed remotely).” (’176 Patent, 4:9-16.)

108. The conversational language processor 120 generates a “domain-specific conversational response” that is “returned to the user as an output.” (’176 patent, 3:60-62.) It may also build “long-term and/or short-term shared knowledge”

in one or more “knowledge source.” (’176 patent, 4:17-19.) “[K]nowledge sources may include information about previous utterances, requests, and other user interactions to inform generating an appropriate response to a current utterance.” (’176 patent, 4:19-22.) “The shared knowledge may include public/non-private (i.e., environmental) knowledge, as well as personal/private (i.e., historical) knowledge.” (’176 patent, 4:22-24.)

109. Figure 2 of the ’176 patent, which I reproduce below, shows an “exemplary advertising system 200.” (’176 patent, 5:10-12.)

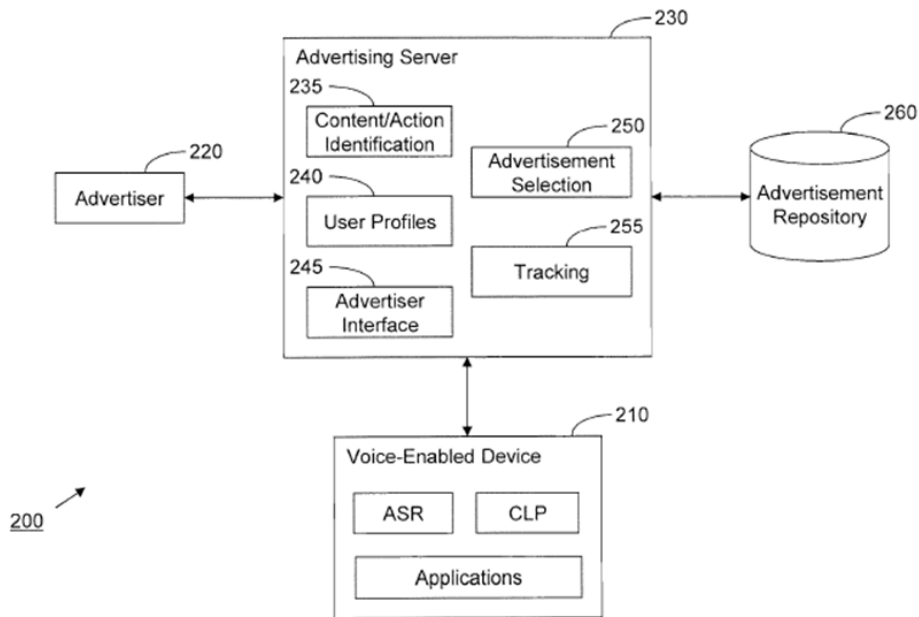


Fig. 2

### **‘176 Patent, Figure 2**

110. System 200 includes an advertising server 230 “for receiving one or more advertisements from an advertiser 220, wherein the advertisements may be

stored in a data repository 260 associated with the server 230.” (’176 patent, 5:12-15.)

111. System 200 includes a voice-enabled device 210 that establishes communications with the advertising server 230. (’176 patent, 5:20-22.) “Subsequently, advertising server 230 may select one or more advertisements from among the advertisements stored in data repository 260, and the selected advertisements may be provided to the voice-enabled device for presentation to a user.” (’176 patent, 5:22-26.)

112. The advertising server 230 of System 200 may include different components, including a content/action identification module 235, a user profile module 240, an advertisement selection module 250, and a tracking module 255. (’176 patent, 5:43-46.)

113. The content/action identification module 235 of the advertising server 230 identifies “content of the voice-based input (e.g., words in the input), requested information (e.g., search results, a web page, music, video, graphics, or other information), requested actions (e.g., calculating a navigation route, placing a telephone call, playing a song, etc.), a category or topic related to the input (e.g., music, business, stocks, sports, navigation, movies, etc.) or other criteria to use in selecting an advertisement.” (’176 patent, 5:56-63.)

114. The user profile module 240 of the advertising server 230 identifies “characteristics of a specific user (e.g., demographics, personal preferences, location-based information, etc.), global user profiles (e.g., demographic profiles, click-through rates, etc.), or other criteria to use in selecting an advertisement.” (’176 patent, 5:64-6:1.)

115. The advertisement selection module 250 of the advertising server 230 “extract[s] relevant information from about [sic] the voice-based input” and “select[s] one or more advertisements relevant to the voice-based input based on information extracted using content/action identification module 235 and/or user profile module 240.” (’176 patent, 5:50-55.) The advertising server 230 communicates the selected advertisements to the voice-enabled device 210. (’176 patent, 6:22-23.)

116. The tracking module 255 of the advertising server 230 tracks the user’s subsequent interactions with an advertisement. (’176 patent, 6:38-40.) “Tracking module 255 may determine whether a conversion or click-through occurs for each advertisement presented to users.” (’176 patent, 6:40-42.) “[T]he tracking information may be used to bill or invoice advertisers 220, as well as to improve subsequent performance and relevance of advertisements selected using advertisement selection module 250.” (’176 patent, 6:59-63.)

## **B. Level of Ordinary Skill in the Art**

117. I understand that certain issues relating to the validity of the '176 patent should be judged from the perspective of a person of ordinary skill in the relevant art, as I discuss below. I have been asked to define the level of a “person of ordinary skill in the art” or “POSITA” at the time the alleged invention as claimed was made. In deciding the level of ordinary skill, I have considered the following factors, which I have been informed are relevant to the determination of the level of ordinary skill in the art.

- levels of education and experience of persons working in the field;
- types of problems encountered in the art;
- prior art solutions to these problems;
- rapidity with which innovations are made; and
- sophistication of the technology.

118. In my opinion, a person of ordinary skill in the art would have a bachelor's degree in computer science, computer engineering, electrical engineering, or 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.

## C. Prosecution and Litigation History

### 1. U.S. Prosecution Before the U.S. Patent & Trademark Office

119. The '176 patent issued from U.S. Patent Application Serial No. 11/671,526, filed February 6, 2007.

120. During prosecution, the Examiner rejected originally filed claims 1-20 as anticipated by U.S. Patent No. 7,461,059 ("Richardson"). (Prosecution History of U.S. Patent No. 7,818,176 ("176 PH"; EX-1002, 152.))

121. Applicants filed a Response on February 17, 2010, amending originally filed independent claims 1 and 11 and adding new claims 21-28. (EX-1002, 182-89.) For example, pending independent claim 1 was amended as follows:

1. (Currently Amended) A method for selecting and presenting advertisements in response to ~~voice-based inputs~~ processing natural language utterances, the method comprising:

receiving ~~a voice-based input, the received input including a~~ natural language utterance containing at least one request at an input device;

recognizing ~~retrieving information from at least one~~ or more words or phrases in knowledge source-based on the received request natural language utterance at a speech recognition engine coupled to the input device;

interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

selecting an advertisement ~~based on~~ in the context established for received request and the retrieved information natural language utterance; and

presenting the selected advertisement ~~to a user~~ via an output device coupled to the conversational language processor.

**'176 Prosecution History, 182**

122. Applicants also argued that Richardson did not anticipate originally filed claims 1-20 because “although Richardson generally describes processing search queries that can be provided in a ‘speech input,’ Richardson does not disclose the specific techniques that amended independent claim 1 recites to process natural language speech inputs.” (EX-1002, 191.) Applicants argued that Richardson “only focuses on the techniques used to search for information from words, letters, or other phrases that can be obtained from speech inputs.” (EX-1002, 191.) In contrast, Applicants argued, “independent claim 1 has been amended to recite various features that recognize words or phrases contained in a natural language utterance and then interpret the recognized words or phrases.” (EX-1002, 191.) Applicants also acknowledged that Richardson generally describes “‘associat[ing] content, such as advertising, with results of a search performed’ on words or phrases in speech inputs,” but they argued that it “does not disclose selecting the advertising content associated with the search ‘in a context’ that a conversational language processor establishes from interpreting the words or phrases in the speech inputs.” (EX-1002, 192.)

123. On May 27, 2010, the Examiner issued a Final Rejection, rejecting pending claims 4, 5, 7, 10, 14, 15, 17, 20, 24, 25, 27, and 28 under § 112. (EX-1002, 205.) The Examiner found that the specification failed to describe the newly added terms “environmental model,” “dictionary and phrase entries that are dynamically

updated,” and “dynamic fuzzy set probabilities.” (EX-1002, 207.) The Examiner also rejected pending claims 1-3 and 11-13 as anticipated by U.S. Patent No. 2010/0063880 (“Atsmon”) (EX-1002, 209-13.) The Examiner also rejected pending claims 4-6, 8, 9, 14-16, 18, and 19 as obvious over Atsmon and U.S. Patent Application Publication 2007/0186165 (“Maislos”) and pending claims 10, 20, 21, and 25 as obvious over Atsmon and U.S. Patent Application Publication 2004/0193420 (“Kennewick”). (EX-1002, 221-25.)

124. Applicants filed a Request for Continued Examination on July 22, 2010, amending the specification “to include material previously incorporated by reference” to address the § 112 rejection. (EX-1002, 239-42, 257-58.) To respond to the § 102 rejection over Atsmon, Applicants cancelled claims 1 and 11, and amended claims 2-3 and 12-13 to depend from claims 22 and 26. (EX-1002, 243-50, 256, 258-59.) Applicants also added new claims 29-54, depending from claims 22 and 26 (EX-1002, 250-55, 259.)

125. The Examiner allowed the pending claims, stating in the Notice of Allowance that the amendments to the specification overcame the § 112 rejections and that claims 22, 23, 26, and 27 being rewritten in independent form made them allowable. (EX-1002, 274.)

## **2. Inter Partes Review before the Patent Trial and Appeal Board**

126. The '176 patent was subject to an IPR petition on two grounds. (“Amazon-VBA-IPR Petition”; EX-1014, 3-4.) The first ground asserted that claims 1-3, 6-19, 22-29, 32-45, and 48-52 were rendered obvious by U.S. Patent Application Publication 2004/0193420 (“Kennewick”), Japan Patent No. 2002-297626A (“Yonebayashi”), and U.S. Patent No. 6,173,250 (“Jong”). (Amazon-VBA-IPR Petition, 16.) The second ground asserted that claims 4, 5, 20, 21, 30, 31, 46, and 47 were rendered obvious by Kennewick, Yonebayashi, Jong, and U.S. Patent No. 7,774,333 (“Colledge”). (EX-1014, 64.).

127. Petitioner argued that, because Kennewick and Jong disclosed a set of rules for mapping phonemes into syllables into words, the “acoustic grammar” limitation was satisfied. (EX-1014, 25-26.) Specifically, Petitioner argued that Jong discloses a device that extracts feature vectors from speech, and a word model device and subword model database that match phonemes to syllables and identify words. (EX-1014, 25.) The PTAB denied institution, stating that the Petitioner failed to advance a construction of acoustic grammar that would allow the PTAB to meaningfully identify whether Kennewick and Jong disclosed what was claimed in the '176 patent. (“Amazon-VBA-IPR Decision”; EX-1015, 18-22.)

128. Petitioner further argued that Kennewick disclosed a system for identifying misrecognition. (EX-1014, 50-52.) Specifically, Petitioner pointed to a

system disclosed in Kennewick wherein there was a “scoring system to determine whether the natural language utterance includes incomplete or ambiguous information.” (EX-1015, 23.) The PTAB denied institution, stating that the Petitioner failed to explain how the scoring system disclosed in Kennewick was analogous to the system disclosed in the ’176 patent. (EX-1015, 24-25.)

### **3. Proceedings in the U.S. District Court**

129. Claims 1, 14, 27, 29, 35, 40, 43, 45, and 46 of the ’176 patent are being asserted against Amazon in a patent infringement action in the U.S. District Court for the District of Delaware. *VB Assets, LLC v. Amazon.com Servs. LLC*, No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019) (“Amazon Litigation”).

130. During those proceedings, certain terms in the ’176 patent were subject to claim construction, where the following constructions were agreed upon or otherwise imposed by the Court:

| <b>Claim Term</b>                                           | <b>Construction</b>                                                                                                                                                                                                                                         |
|-------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| “model”<br><br>‘176 patent, claims 2-5, 18-21, 28-31, 44-47 | “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system” (“Amazon Litigation Mem. Op. Regarding Claim Construction”; EX-1011, 1) |

|                                                                                                             |                                                                                                                                                                                                                                                                                                                          |
|-------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| “acoustic grammar”<br>‘176 patent, claims 1, 16, 27, 42                                                     | “grammar of phonotactic rules of the English language that maps phonemes to syllables” (EX-1011, 2)                                                                                                                                                                                                                      |
| “speech recognition engine” / “speech recognition”<br>‘176 patent, claims 1, 10, 14, 25, 27, 36, 40, 42, 51 | “software or hardware that recognizes the words or phrases in the natural language utterance” (EX-1011, 2)                                                                                                                                                                                                               |
| “context”<br>‘176 patent, claims 1, 14, 27, 40                                                              | The Court found that no construction was necessary for this term. (Amazon Litigation Mem. Op. Regarding Claim Construction, 2)<br>Amazon had proposed that the term should mean “subject domain(s) that applies to a user utterance.” (“Amazon Litigation Second Am. Joint Claim Construction Chart”; EX-1012, EX. A-7.) |

131. I understand that during trial in the Amazon Litigation, Mike Kennewick, one of the inventors of the ’176 patent testified that he did not invent the concept of a “conversational language processor.” (“Amazon Litigation Trial Tr. Vol. 1”; EX-1013, 226:25-227:5.)

#### **D. Claim Construction**

132. I understand that “claim construction” is the process of determining a patent claim’s meaning. I understand that the Patent Office applies the same claim construction standard used in district courts which seeks to give claim terms their plain and ordinary meaning to a POSITA at the time of the claimed invention in view

of the claims, specification, and prosecution history. I also understand that if the patent applicant gave a term a special meaning in the specification or during prosecution, then the special meaning should be used.

133. As discussed above, I understand that in the Amazon Litigation and the Amazon and PO took certain claim construction positions that were eventually ruled on. In addition, I understand that in the Amazon-VBA-IPR the Board interpreted “acoustic grammar” as a “collection of the phonemes, or distinct units of sound of a spoken language, linked together to form syllables, which are linked together to form the words of the language.” (EX-1015, 10-12.)

134. For purposes of my analysis here, I have applied the meaning of the terms that the district court has adopted or which the parties agreed as well as the plain and ordinary meaning as it would be understood by a POSITA.

## **V. GROUND 1: THE COMBINATION OF TASCHEREAU AND HANSEN RENDERS OBVIOUS CLAIMS 1-2, 9-11, 27-28, 35-37**

### **A. Overview of the Combination**

135. The combination of WO 2004/055781 to Taschereau, et al. (“Taschereau”; EX-1005) and U.S. Patent 5,640,490 to Hansen et al. (“Hansen”; EX-1006) discloses every limitation of claims 1-2, 9-11, 27-28, and 35-37. Taschereau and Hansen are analogous art to the ’176 patent because both references are directed to speech recognition systems and software.

136. I describe both Taschereau and Hansen further below.

## **1. Taschereau**

137. Taschereau discloses a speech recognition and response system for use in, for example, telephone directory assistance services. (Taschereau, Abstract.) The disclosed system receives and analyzes speech using an automatic speech recognition (ASR) system which generates a “word list” and passes a preliminary interpretation through a “dynamically generated grammar” for further analysis. (Taschereau, 7:24-16:3; Fig. 11.) As noted in Taschereau:

In the prior art, ASR is a “one pass” process: a grammar is generated, applied and the result is examined. The process according to the invention is a “multi pass” process: a grammar is generated which is designed to result in the generation of a one or more “latent grammars.”

(Taschereau, 13:8-10.)

138. An utterance from a user (for example, a “request for Business Name”), matches the words contained within the utterance to a grammar list, and provides an oral response to the user. (Taschereau, Abstract, 13:26-14:3.) Taschereau further discloses that the system can use the user’s utterance to determine and provide targeted advertisements to the user, for example, a competitor business. (Taschereau, 25:5-11; claim 9.)

139. The system disclosed by Taschereau includes at least four steps: transformation, word map, grammar generation and grammar interpretation. (Taschereau, 7:24-8:1.)

140. In the “transformation” step, an “item to be represented” (for example, a list of businesses in a telephone directory assistance embodiment), is transformed to “create a Word List appropriate to the grammar.” (Taschereau, 8:5-14.) An exemplary word list is shown in Figure 3:

**FIG. 3**

**Transformations**

| NAME                                | TERMS                                                 |
|-------------------------------------|-------------------------------------------------------|
| A & A Piano Service Ltd             | a, and, a, double, piano, service, limited, l-t-d     |
| A & A Satellite Express Ltd         | a, and, a, double, satellite, express, limited, l-t-d |
| A-1 Aberdeen Piano Tuning & Repairs | a, one, aberdeen, piano, tuning, and, repairs         |
| A-White Rock Roofing                | a, white, rock, roofing                               |
| Bryans Auto Repair                  | bryans, auto, repair                                  |
| National Satellite Communications   | national, satellite, communications                   |
| North Bluff Auto Services           | north, bluff, auto, services                          |
| Priority Roofing Ltd                | priority, roofing, limited, l-t-d                     |
| Thomson Roofing Service             | thomson, roofing, service                             |
| White Rock Automotive Services Ltd  | white, rock, automotive, services, limited, l-t-d     |

**Taschereau, Figure 3**

141. “The Word Map is a list of terms ... and corresponding references to the item to be represented. Each entry in the Word Map maps at least a single term and a reference to an item to be represented.” (Taschereau, 9:15-19.) The Word Map may include a Word Base that contains the base term of a given word (“[f]or example, repairing, repaired and repair have the same base term: repair”). (Taschereau, 9:26-29.) A portion of an exemplary Word Map is shown in Figure 4:

**FIG 4****Word Map**

| Word      | Word Base | Reference     | Use Count |
|-----------|-----------|---------------|-----------|
| a         | a         | TV12401305638 | 6         |
| and       | and       | TV12401305638 | 3         |
| a         | a         | TV12401305638 | 6         |
| double    | double    | TV12401305638 | 2         |
| piano     | piano     | TV12401305638 | 2         |
| service   | service   | TV12401305638 | 2         |
| limited   | limited   | TV12401305638 | 4         |
| l-t-d     | l-t-d     | TV12401305638 | 4         |
| a         | a         | TV12401305471 | 6         |
| and       | and       | TV12401305471 | 3         |
| a         | a         | TV12401305471 | 6         |
| double    | double    | TV12401305471 | 2         |
| satellite | satellite | TV12401305471 | 2         |
| express   | express   | TV12401305471 | 1         |
| limited   | limited   | TV12401305471 | 4         |
| l-t-d     | l-t-d     | TV12401305471 | 4         |
| a         | a         | TV12401305135 | 6         |
| one       | one       | TV12401305135 | 1         |
| aberdeen  | aberdeen  | TV12401305135 | 1         |
| piano     | piano     | TV12401305135 | 2         |
| tuning    | tune      | TV12401305135 | 1         |
| and       | and       | TV12401305135 | 3         |
| repairs   | repair    | TV12401305135 | 1         |

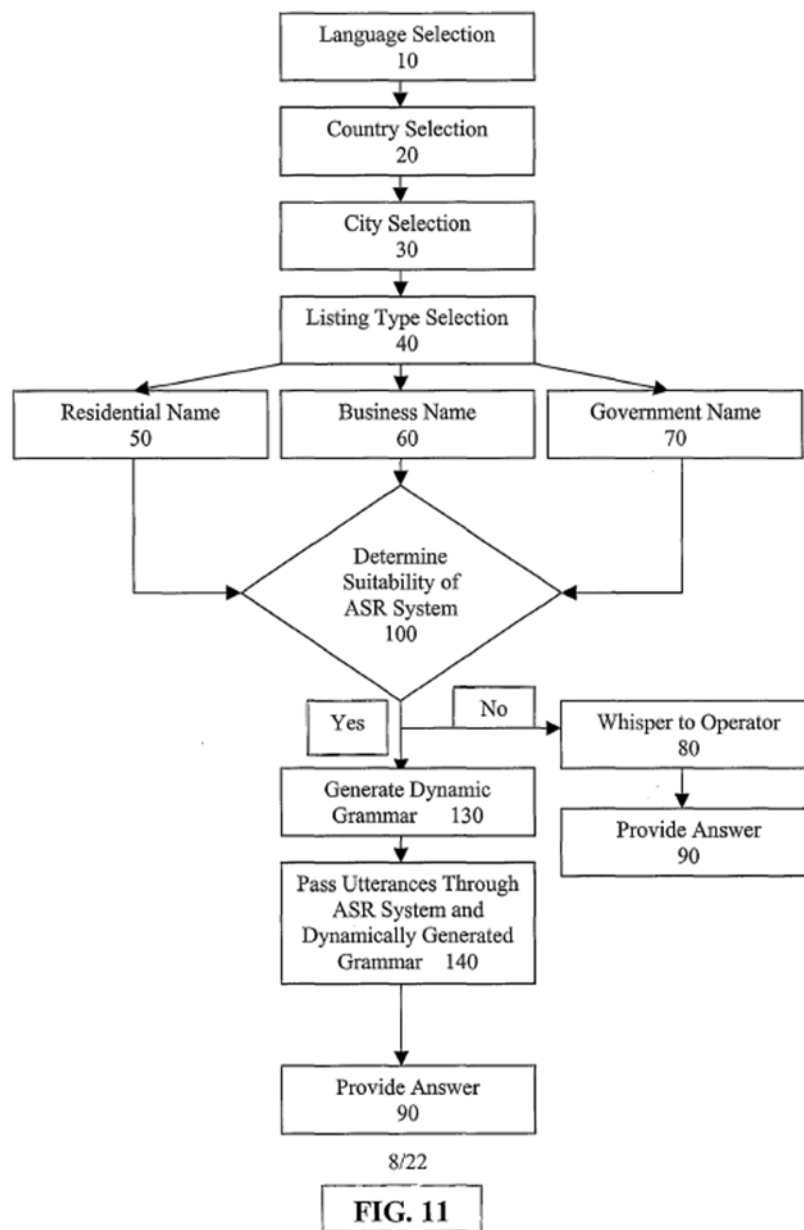
**Taschereau, Figure 4**

142. “An objective of the grammar generation process [step three] is to generate a single list of terms which can be used in a subsequent process to determine which items to be represented are being referenced while keeping the number of terms used in the grammar to a number suitable for practical application.” (Taschereau, 10:21-24.) “The process commences by generating a list which contains all of the distinct terms from the Word Map, called a ‘Word List.’ If the number of items in the list is unsuitable for practical application (i.e. it is too large), the list is ‘trimmed.’ The ‘trimming’ process removes words based on usage

frequency and other criteria from the list.” “The ‘trimming’ process removes words based on usage frequency and other criteria from the list.” (Taschereau, 10:24-29.) “Once a final trimmed Word List has been determined, it is assembled into an ASR grammar following common practices.” (Taschereau, 12:20-21.)

143. To interpret the grammar against an utterance, the grammar must first be evaluated based on the dynamic “associated return result frequency and confidence scores.” (Taschereau, 13:21-14:6.) The Word List entries deemed the best are placed into a “latent grammar.” (Taschereau, 14:8-9.) “The utterance is then run through the latent grammar (more than once if necessary) and an answer is provided.” (Taschereau, 17:11-12.)

144. “Figure 11 is a flow chart showing a voice recognition system using the described invention.” (Taschereau, 6:1-2.)



**Taschereau, Figure 11**

145. “The making of a request for a business also provides an opportunity to target an advertisement. For example when a request is made for a restaurant in a certain geographic area, a competitor could present an advertisement with an inducement (e.g. a coupon or the like) in an attempt to lure that customer to a

different establishment. By using the information available about the user and the listing the user is looking for, very precise advertisements can be presented to the user.” (Taschereau, 25:5-11.)

## **2. Hansen**

146. While Taschereau discloses receiving a natural language utterance, recognizing one or more words or phrases in the natural language utterance, interpreting the natural language utterance, including establishing a context for the natural language utterance, and providing an oral response, it does not explicitly disclose all details for how the disclosed system performs such steps. A POSITA would have looked to known ways of implementing such a system. One well known way of conducting such analysis is to map a stream of phonemes contained in the natural language utterance as taught in, for example, Hansen.

147. Hansen discloses systems for “accurately recognizing continuous human speech” during “speech recognition” (Hansen, 1:8-10) by “identifying the phoneme sound types that are contained within an audio speech signal” (Hansen, Abstract). As explained by Hansen:

Simplistically, speech may be considered as a sequence of sounds taken from a set of forty or so basic sounds called “phonemes.” Different sounds, or phonemes, are produced by varying the shape of the vocal tract through muscular control of the speech articulators (lips, tongue, jaw, etc.). A stream of a particular set of phonemes will collectively represent a word or a phrase. Thus, extraction of the

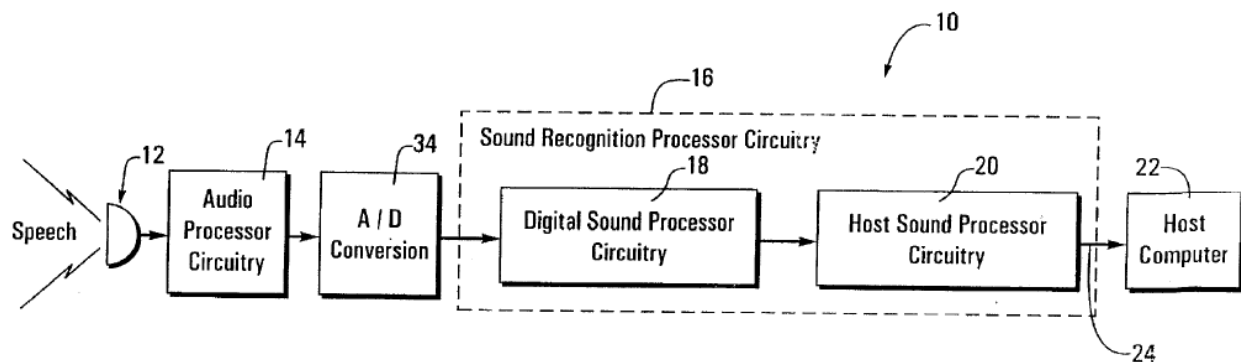
particular phonemes contained within a speech signal is necessary to achieve voice recognition.

(Hansen, 1:24-33.)

148. The disclosed “object[s]” of the invention include “a speech recognition system and method that is capable of accurately extracting various sound characteristics from a speech signal, and then converting those sound characteristics into representative phonemes” and “a speech recognition system and method that is capable of converting a stream of phonemes into an intelligible format.” (Hansen, 2:28-35.)

149. “In addition to extracting phoneme sound types from the incoming audio speech signal, the sound recognition means implements various linguistic processing techniques to translate the phoneme string into a corresponding word or phrase.” (Hansen, 3:7-11.)

150. The process for mapping a stream of phonemes from an utterance and creating a preliminary interpretation is disclosed in Figure 1, reproduced below.



**Hansen, Figure 1**

151. “The system 10 includes an audio processor means for receiving an audio speech signal and for converting that signal into a representative audio electrical signal.” (Hansen, 4:36-38.) The audio processor circuitry 14 “receives the audio electrical signal generated by the microphone 12, and then functions so as to condition the signal so that it is in a suitable electrical condition for digital sampling.” (Hansen, 4:48-51.) A/D conversion circuitry 34 “receives the audio electrical signal, which is in an analog format, and converts it to a digital format, outputting a digitized audio signal.” (Hansen, 4:55-57.) Sound recognition processor circuitry 16 “programmably analyzes the digitized version of the audio signal in a manner so that it can extract various acoustical characteristics from the signal. Once the necessary characteristics are obtained, the circuitry 16 can identify the specific phoneme sound types contained within the audio speech signal.” (Hansen, 4:61-67.) “As the processor circuitry 16 translates the phoneme string, the corresponding word or phrase is passed to the host computer 22.” (Hansen, 5:47-49.)

152. “Once the sound recognition processor circuitry 16 has extracted the corresponding phoneme sounds, it programmably invokes a series of linguistic program tools. In this way, the processor circuitry 16 translates the series of identified phonemes into the corresponding syllable, word or phrase.” (Hansen, 5:30-35.) A “Linguistic processor receives the data structures, and translates the sound stream (i.e., stream of phonemes) into the corresponding English letter,

syllable, word or phrase. This translation is generally accomplished by performing a variety of linguistic processing functions that match the phonemic sequences against entries in the system lexicon.” (Hansen, 10:2-8.) “The presently preferred linguistic functions include a phonetic dictionary look-up, a context checking function and database, and a basic grammar checking function.” (Hansen, 10:8-11.)

153. An exemplary process used by the Linguistic Processor is shown in Figure 10, reproduced below.

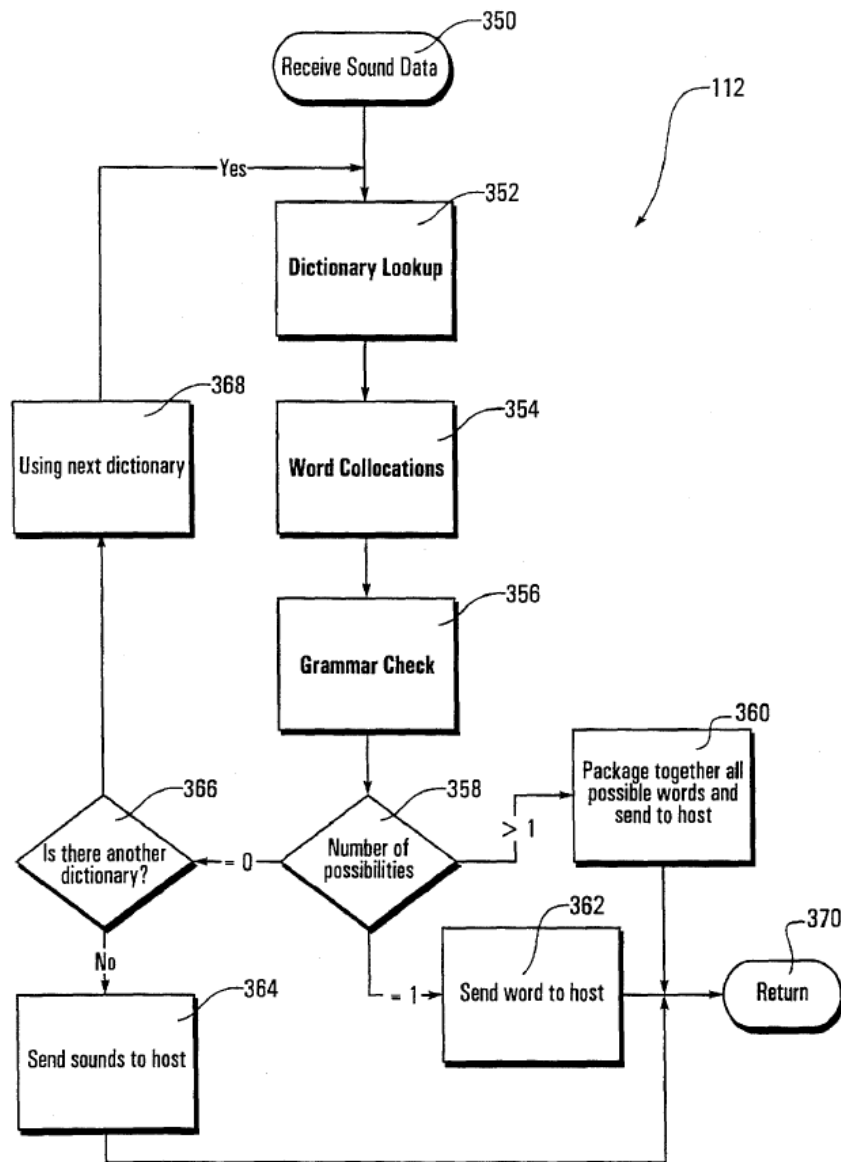


FIG. 10

### Hansen, Figure 10

154. "[T]he Dictionary Lookup function utilizes a phonetic-English dictionary that contains the English spelling of a word along with its corresponding phonetic representation. The dictionary can thus be used to identify the English word

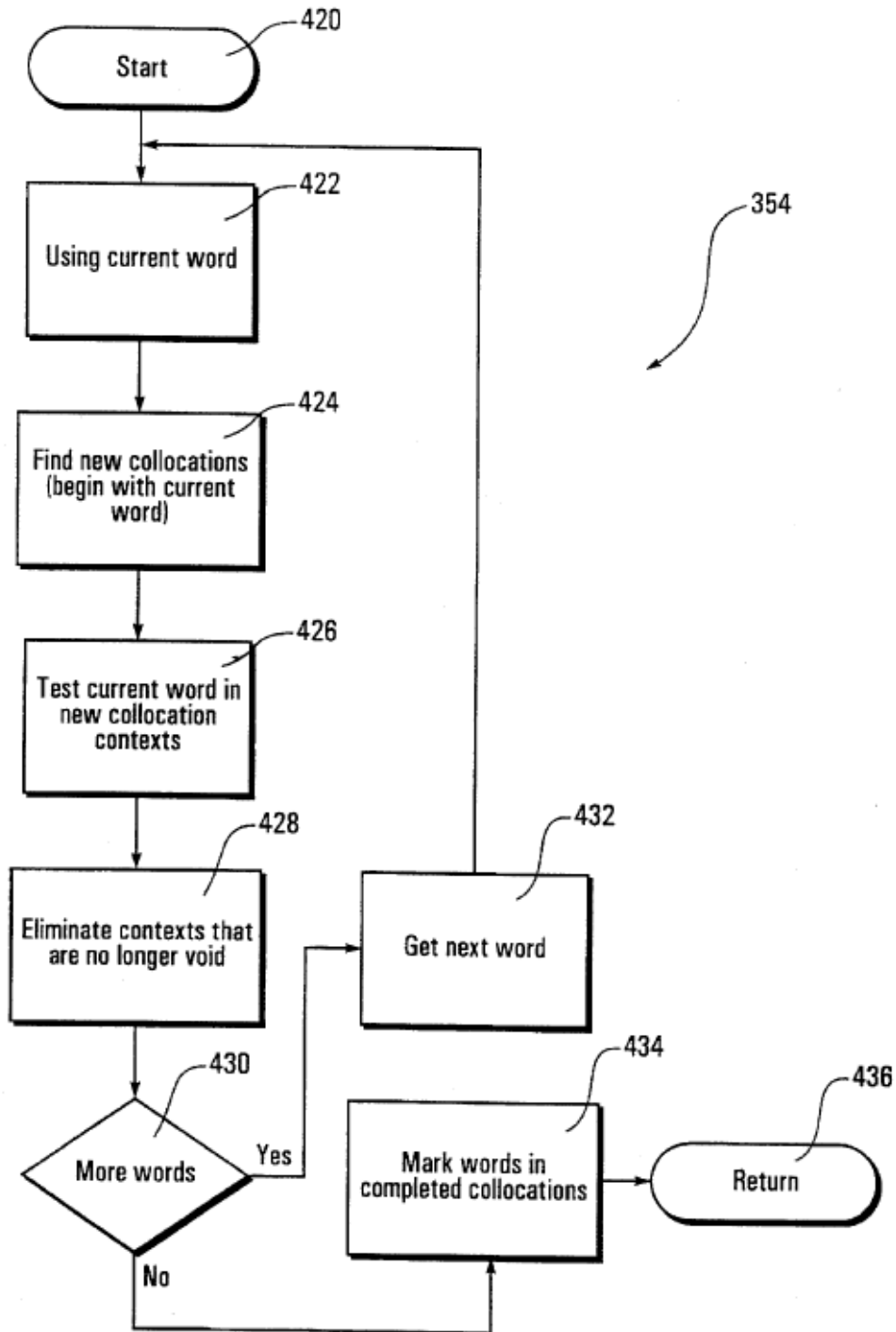
that corresponds to a particular stream of phonemes.” (Hansen, 24:39-44.)

“Dictionary Lookup scans the appropriate dictionary to determine if the incoming sequence of sounds ... form a complete word, or the beginnings of a possible word.” (Hansen, 24:62-65; Fig. 10A.)

155. “[A]s each phoneme sound is received, it is added to the end of a[sic] all non-completed ‘sequences.’ Each sequence is compared to the contents of the dictionary to determine if it leads to a possible word. When a valid word (or set of possible words) is identified, it is passed to the next functional block within the Linguistic Processor portion of the program for further analysis.” (Hansen, 25:1-7.)

This block (354) is called “word collocations.” (Hansen, Fig. 10.)

156. “[T]he Word Collocations function monitors the word possibilities that have been identified by the Dictionary Lookup function to see if they form a ‘common’ word collocation.” (Hansen, 26:20-23.) The word collocations analysis process is shown in Figure 10B, reproduced below.



**Hansen, Figure 10B**

157. “Beginning at program step 420, a set of word possibilities are received. Beginning with one of those words at step 422, the host sound processor 54 next

proceeds to program step 424 where it obtains any collocation(s) that have been formed by preceding words. The existence of such collocations would be determined by continuously comparing words and phrases to the collocation dictionary contents. If such a collocation or collocations exist, then the current word possibility is tested to see if it fits within the collocation context. At step 428, those collocations which no longer apply are discarded.” (Hansen, 26:32-42.)

158. “Another optional function that can be used to resolve remaining word possibilities is the ‘Grammar Check’ function, shown at block 356 in FIG. 10. This function evaluates a word possibility by applying certain grammatical rules, and then determining whether the word complies with those rules. Words that do not grammatically fit can be eliminated as possibilities, or assigned lower confidence levels.” (Hansen, 26:53-60.)

159. “If at program step 358 more than one word possibility still remains, the remaining words are all sent to the Command Processor. Similarly, if only one word possibility remains, that word is sent to the directly to the Command Processor portion of the algorithm.” (Hansen, 27:27-31.) “[T]he Command Processor is a series of program steps that are executed by a Host Computer 22.” (Hansen, 27:38-41.) “The Command Processor [] receives each word, and determines the manner by which it should be used on the host computer 22. For example, a spoken word may be input as text directly into an application, such as a wordprocessor document.

Conversely, the spoken word may be passed as a command to the operating system or application.” (Hansen, 27:44-49.) “If the incoming word is neither a[n] operating system command, or an application command, then program step 460 is executed, where it is determined whether the Command Processor is still in a ‘command mode.’ If so, the word is discarded at step 464, and essentially ignored. However, if the Command Processor is not in a command mode, then the word will be sent directly to the current application as text.” (Hansen, 28:4-11.)

### **3. Motivation to Combine**

160. A POSITA would have been motivated to combine the teachings of Hansen regarding the specific details of speech recognition, including mapping phonemes to syllables phonemically represented in an acoustic grammar, with the speech recognition and response system disclosed in Taschereau.

161. A POSITA would have been motivated to make this combination for numerous reasons. First, a POSITA would have been motivated to make the above combination because Taschereau is silent regarding the details of how its ASR system converts speech to text. Therefore, a POSITA seeking to build a system like that disclosed in Taschereau would have necessarily needed to seek out information detailing how the ASR system recognizes and processes utterances. A POSITA would have therefore been motivated to search for references providing details on speech recognition processing, particularly associated with how to recognize and

process utterances spoken by users in the Taschereau system. A POSITA would have been led to Hansen which discloses well known speech recognition processing techniques.

162. Second, a POSITA would have also been motivated to make the above combination to improve the speech recognition system, a motivation expressly stated in Taschereau. Taschereau teaches that “[t]he ASR system can also be improved through additional audio processing in addition to or in place of gain control, for example by examining and adjusting for attributes particular to the utterance to be recognized ....” (Taschereau, 18:6-9.) Taschereau further recognizes that “[v]arious complex process[sic] which analyse the utterance and remove portions which would hinder the ASR recognition” are “[e]xample[s] of audio processing which may be applied.” (Taschereau, 18:10-25.) Hansen describes such a process which may be used in audio processing to analyze the utterance. Further, it was well-known long before the ’176 patent that mapping phonemes to syllables phonemically represented in an acoustic grammar. (*See, e.g.*, U.S. Patent No. 6,804,330 (“Jones ’330”; EX-1017), 7:56-8:63 (describing process for mapping a stream of phonemes to facilitate speech recognition).)

163. Finally, the combination is nothing more than the application of a known technique (mapping phonemes to syllables phonemically represented in an acoustic grammar as disclosed in Hansen) to a similar known system (Taschereau’s

ASR system) that operates in the same way (analyzing utterances through investigation of specific words and dynamic development of grammars). The system of Taschereau is ready for improvement and the simple substitution of one known element (Hansen ASR using phoneme mapping) for another (Taschereau's agnostic ASR). The results of the combination would have been predictable as I discuss below. And, a POSITA would have been motivated to make the combinations for the reasons discussed above.

164. A POSITA would have had a reasonable expectation of success in the combination and the results of the combination would have been predictable. Combining Taschereau with Hansen would merely integrate one component (mapping phonemes to syllables phonemically represented in an acoustic grammar as disclosed in Hansen) into Taschereau's system of receiving and responding to user's utterances. Taschereau expressly states that its ASR system can be improved through additional audio processing techniques. (Taschereau, 18:6-9.) Hansen provides an example of a well-known processing technique that can be used in the Taschereau system. (*See Hansen, 7:56-62.*) Integrating Hansen's speech recognition and processing into Taschereau's system to improve the processing of utterances would have been well within the skill set of a POSITA before the earliest priority date of the '176 patent.

165. Taschereau and Hansen are both in the same field of the '176 patent—using speech recognition to facilitate access to information (e.g., data or files) or programs. (*See* '176 patent, 1:66-2:2 (“a system and method for selecting and presenting advertisements based on natural language processing of voice-based inputs is provided.”); Taschereau, 1:6-8 (“This invention relates to systems and methods of voice recognition, and more particularly voice recognition used in the context of directory assistance”), 4:25-27 (“The method and processes described herein implement technologies for ASR systems that are especially useful in applications where the possible utterances represent a large or very large collection of possibilities (i.e. a large grammar is required).”); Hansen, 27:53-65 (“program step 450 causes the host computer 22 to receive a word created by the Linguistic Processor portion of the algorithm. The host computer 22 then determines, at step 452, whether the word received is an operating system command .... If the incoming word does not constitute an operating system command, step 456 is executed, where it is determined if the word is instead an application command, as for instance a command to a wordprocessor or spreadsheet”); 28:25-30 (“the system and method of the current invention act as a means for entering data and/or commands to a standard personal computer”).)

## B. Independent Claims 1 and 27

166. Independent claims 1 and 27 include substantially overlapping processing limitations ([1A]-[1G]/[27A]-[27G]), as shown in the table below. The only substantive difference between the claims is that claim 1 claims a method for selecting and presenting advertisements and claim 27 claims a system that practices that same method for selecting and presenting advertisements. Because of the overlap the limitations in these claims, I address these limitations together. In this table, I also provide a citation to the section discussing each limitation:

| Claim 1                                                                                                                                                                                                                          | Claim 27                                                                                                                                                                                                                                                                                                     | Section |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| [1P] A method for selecting and presenting advertisements in response to processing natural language utterances, comprising:                                                                                                     | [27P] A system for selecting and presenting advertisements in response to processing natural language utterances, comprising                                                                                                                                                                                 | §V.B.1  |
| [1A] receiving a natural language utterance containing at least one request at an input device;                                                                                                                                  | [27A] an input device that receives a natural language utterance containing at least one request at an input device;                                                                                                                                                                                         | §V.B.2  |
| [1B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or phrases in the natural language utterance includes: | [27B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to: | §V.B.3  |

| <b>Claim 1</b>                                                                                                                                                                                                                                              | <b>Claim 27</b>                                                                                                                                                                                                                                                                                                         | <b>Section</b> |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| [1C] mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and                                                                                        | [27C] map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and                                                                                                                                                       | §V.B.4         |
| [1D] generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;                   | [27D] generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;                                                                                | §V.B.5         |
| [1E] interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | [27E] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | §V.B.6         |
| [1F] selecting an advertisement in the context established for the natural language utterance; and                                                                                                                                                          | [27F] select an advertisement in the context established for the natural language utterance; and                                                                                                                                                                                                                        | §V.B.7         |
| [1G] presenting the selected advertisement via an output device coupled to the conversational language processor.                                                                                                                                           | [27G] present the selected advertisement via an output device.                                                                                                                                                                                                                                                          | §V.B.8         |

## 1. Preamble [1P] / Preamble [27P]

[1P] A method for selecting and presenting advertisements in response to processing natural language utterances, comprising:

[27P] A system for selecting and presenting advertisements in response to processing natural language utterances, comprising

167. To the extent the preamble is limiting, the combination of Taschereau and Hansen discloses “[a] method for selecting and presenting advertisements in response to processing natural language utterances” as well as “[a] system for selecting and presenting advertisements in response to processing natural language utterances,” as discussed below in §§V.B.1-8. Taschereau discloses a system that processes natural language utterances and teaches that “[t]here are several opportunities for advertisements to be presented to a user during the automation process as described.” (Taschereau, 24:25-26.) As discussed below, the combination of Taschereau and Hansen discloses a “system” that practices a “method” “for selecting and presenting advertisements in response to processing natural language utterances,” so I analyze these claims together.

168. As one example, Taschereau describes that after the user makes a natural language utterance, there is an opportunity to target an advertisement: “The making of a request for a business also provides an opportunity to target an advertisement.” (Taschereau, 25:5.) Taschereau describes that the system can provide a targeted advertisement based on the information in the user’s “natural

*language utterance[]*”: “By using the information available about the user and the listing the user is looking for, very precise advertisements can be presented to the user.” (Taschereau, 25:10-11.) “For example when a request is made for a restaurant in a certain geographic area, a competitor could present an advertisement with an inducement (e.g. a coupon or the like) in an attempt to lure that customer to a different establishment.” (Taschereau, 25:5-11; *see also* claim 9.)

169. Taschereau also claims this method in claim 9:

A method of providing directory assistance to a user comprising:

(a) receiving an utterance from a user;

**(b) determining a listing in response to said utterance;**

**(c) providing an advertisement to said user before providing said listing to said user;**

wherein said user is not charged an additional fee for the directory assistance.

(Taschereau, claim 9.)

170. The combination of Taschereau and Hansen discloses an “[a] *method for selecting and presenting advertisements in response to processing natural language utterances*” and “[a] *system for selecting and presenting advertisements in response to processing natural language utterances.*”

## **2. Receiving Natural Language Request [1A] / [27A]**

[1A] **receiving a natural language utterance containing at least one request at an input device;**

[27A] **an input device that receives a natural language utterance containing at least one request at an input device;**

171. The combination of Taschereau and Hansen discloses “*receiving a natural language utterance containing at least one request at an input device*” and “*an input device that receives a natural language utterance containing at least one request at an input device.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

172. The ’176 patent describes a “request” as “performing an action, retrieving information, or various combinations thereof.” (’176 patent, 4:1-3.) The ’176 patent also provides examples of types of “requests”:

For example, content/action identification module 235 may identify content of the voice-based input (e.g., words in the input), requested information (e.g., search results, a web page, music, video, graphics, or other information), requested actions (e.g., calculating a navigation route, placing a telephone call, playing a song, etc.), a category or topic related to the input (e.g., music, business, stocks, sports, navigation, movies, etc.), or other criteria to use in selecting an advertisement.

(’176 patent, 5:56-6:1.)

173. Taschereau discloses a system that receives natural language utterances containing at least one request, like those described in the ’176 patent. Taschereau discloses the process for receiving and storing the natural language utterance: “[t]he process is facilitated by recording or saving the original utterance of the user as applied to the initial or first grammar and applying the same utterance to subsequent

grammars which are dynamically generated (or may have been previously generated).” (Taschereau, 6:26-7:1.)

174. Taschereau also discloses that the user’s natural language utterance can include a request. (Taschereau, 23:19-21 (“The system and method described above can also serve to direct services to users or direct users to services. For example when a user requests the phone number of a taxi company, it is likely that user is actually trying to have a taxi sent to a particular location.”).) For example, Taschereau discloses the following types of requests a user can make with a natural language utterance:

- “For the purposes of an example, let us assume that the user responded to a request for Business Name with ‘Kearney Funeral Home’.” (Taschereau, 14:1-3);
- “In a voice recognition system according to the invention, one of the primary goals is to create a transparent interface, such that every time a requestor calls for assistance, whether the request is handled by voice recognition or by a human operator, the same pattern of questions will be provided in the same order.” (Taschereau, 16:5-8);
- “Another feature of the grammar based on the word list can be weighting the grammar towards more frequently requested listings (‘FRLs’). Certain listings are more frequently requested, such as taxis,

pizza restaurants, hotels and tourist destinations.” (Taschereau, 21:14-16.)

175. As shown in the examples above, one purpose of calling the directory assistance system described by Taschereau is to request information about a business.

176. Taschereau also discloses receiving the natural language utterance request “*at an input device*.” Taschereau discloses that “a user from a phone booth is most likely to be sent to the ASR system.” (Taschereau, 20:24-26.) Here, the microphone in the handset is an “*input device*” that receives the natural language utterance containing a request.

177. Taschereau provides additional examples of input devices that can be used to receive a natural language utterance from a user:

The method and processes described herein are described with respect to telephone directory assistance systems although the process is not limited to such application and can be used in situations wherever voice recognition is used, including mobile phone interfaces, in-vehicle systems, and the like.

(Taschereau, 5:3-6.)

178. Therefore, Taschereau discloses input devices such as landline phones, mobile phones, in-vehicle systems and other systems where voice recognition is used. (*See* Taschereau, 5:3-6.)

179. The combination of Taschereau and Hansen discloses an “*receiving a natural language utterance containing at least one request at an input device.*”

### 3. Recognizing One or More Words in the Utterance [1B] / [27B]

[1B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or phrases in the natural language utterance includes:

[27B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to:

180. The combination of Taschereau and Hansen discloses “*recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or phrases in the natural language utterance includes*” and “*a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

181. Taschereau includes a “*speech recognition engine coupled to [an] input device,*” the “*input device*” being the ASR recognizer. Taschereau discloses that a user’s “speech is simultaneously ‘**recognized**’ and ‘recorded’.” (Taschereau, 13:12-

15.) The ASR recognizer (the “*input device*”) is therefore coupled to a speech recognition engine that obtains the information from the ASR recognizer. To obtain the information from the ASR recognizer, the speech recognition engine must necessarily be coupled to it.

182. The ASR system of Taschereau receives a natural language utterance and employs four steps (transformation, word map, grammar generation and grammar interpretation) in order to recognize one or more words or phrases in the natural language utterance. (Taschereau, 7:24-8:1.)

183. Claim 1 of Taschereau claims using the disclosed system to recognize the words in the user’s utterance:

1. A method of matching an utterance comprising a word to a record in a database using an automated speech recognition system comprising:
  - (a) forming a word list comprising a selection of words from said records in said database;
  - (b) using the automated speech recognition system to determine the best possible matches of the word in said utterance to the words in said word list;
  - (c) creating a grammar of records in said database that contain at least one of said best possible matches;
  - (d) using the automated speech recognition system to match said utterance to a record within said grammar.

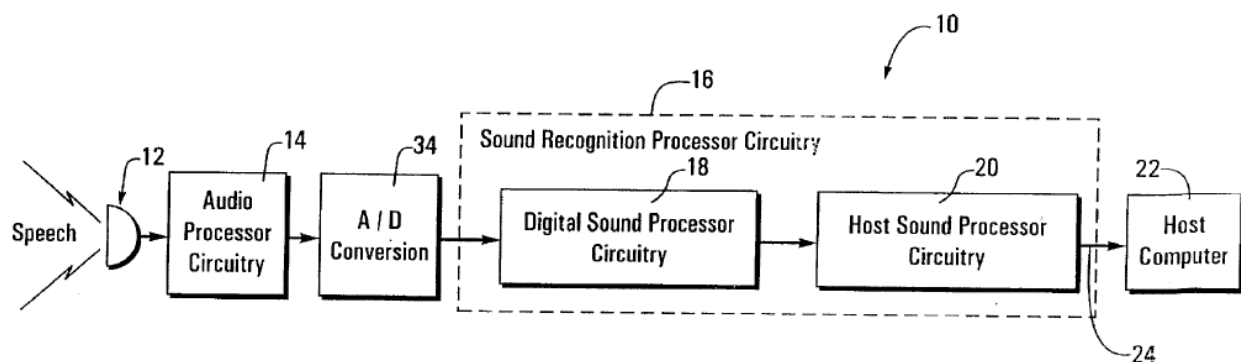
(Taschereau, claim 1.) Taschereau thus discloses a “*speech recognition engine*” under the Amazon-VBA-Litigation construction.

184. The phone handset (“*input device*”) is “*coupled*” to a speech recognition engine. (Taschereau, 20:21-26 (noting that a phone booth user of the

ASR system “may have a longer [] time to arrive at the answer” as compared to a phone operator).)

185. Hansen likewise discloses a “*speech recognition engine coupled to [an] input device.*” As noted above, one of the disclosed “object[s]” of the invention in Hansen is “a speech recognition system and method that is capable of accurately extracting various sound characteristics from a speech signal, and then converting those sound characteristics into representative phonemes.” (Hansen, 2:28-35.) “Once the sound recognition processor circuitry 16 has extracted the corresponding phoneme sounds, it programmably invokes a series of linguistic program tools. In this way, the processor circuitry 16 translates the series of identified phonemes into the corresponding syllable, word or phrase.” (Hansen, 5:30-35.) Hansen thus discloses a “*speech recognition engine*” under the Amazon-VBA-Litigation construction.

186. The speech recognition process in Hansen is depicted in Figure 1, reproduced below.



**Hansen, Figure 1**

187. “The system 10 includes an audio processor means for receiving an audio speech signal and for converting that signal into a representative audio electrical signal.” (Hansen, 4:36-39.) The audio processor circuitry 14 “receives the audio electrical signal generated by the microphone 12, and then functions so as to condition the signal so that it is in a suitable electrical condition for digital sampling.” (Hansen, 4:48-51.) The disclosed microphone 12 is an “*input device*.”

188. “This digitized audio signal is then passed to a sound recognition means, which in the preferred embodiment corresponds to the block designated at 16 and referred to as the sound recognition processor circuitry. Generally, the sound recognition processor circuitry 16 programmably analyzes the digitized version of the audio signal in a manner so that it can extract various acoustical characteristics from the signal. Once the necessary characteristics are obtained, the circuitry 16 can identify the specific phoneme sound types contained within the audio speech signal.” (Hansen, 4:58-67.) “Once the sound recognition processor circuitry 16 has extracted

the corresponding phoneme sounds, it programmably invokes a series of linguistic program tools. In this way, the processor circuitry 16 translates the series of identified phonemes into the corresponding syllable, word or phrase.” (Hansen, 5:30-35.) Hansen thus discloses a “*speech recognition engine*” under the Amazon-VBA-Litigation construction.

189. The combination of Taschereau and Hansen discloses an “*recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device wherein recognizing the words or phrases in the natural language utterance includes*” and “*a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to.*”

#### **4. Mapping Phonemes [1C] / [27C]**

[1C] mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

[27C] map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

190. The combination of Taschereau and Hansen discloses “*mapping a stream of phonemes contained in the natural language utterance to one or more*

*syllables that are phonemically represented in an acoustic grammar” and “map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar.”* These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

191. As discussed above, I understand that in the prior District of Delaware proceeding the term “acoustic grammar” was construed to mean “grammar of phonotactic rules of the English language that maps phonemes to syllables.” (Amazon Litigation Mem. Op. Regarding Claim Construction, 2.)

192. I further understand that in the prior IPR proceeding the Board interpreted acoustic grammar as follows:

Thus, Kennewick ’409 explains that an “acoustic grammar” is a collection of the phonemes, or distinct units of sound of a spoken language, linked together to form syllables, which are linked together to form the words of the language. On this record and for the purposes of this Decision, we interpret “acoustic grammar” as used in the ’176 patent in the same manner.

(Amazon-VBA-IPR Decision, 12.)

It is my opinion that this definition of “acoustic grammar” is consistent with the definition used in the District of Delaware proceeding. I have applied them to my opinions here.

193. Hansen discloses “a speech recognition system and method that is capable of converting a stream of phonemes into an intelligible format.” (Hansen, 2:33-35.)

194. Sound recognition processor circuitry 16 “programmably analyzes the digitized version of the audio signal in a manner so that it can extract various acoustical characteristics from the signal. Once the necessary characteristics are obtained, the circuitry 16 can identify the specific phoneme sound types contained within the audio speech signal.” (Hansen, 4:62-67.)

195. Hansen discloses that a stream of phonemes is mapped to syllables using a phonetic dictionary. “Once the sound recognition processor circuitry 16 has extracted the corresponding phoneme sounds, it programmably invokes a series of linguistic program tools. In this way, the processor circuitry 16 translates the series of identified phonemes into the corresponding syllable, word or phrase.” (Hansen, 5:30-35.) A “Linguistic processor receives the data structures, and translates the sound stream (i.e., stream of phonemes) into the corresponding English letter, syllable, word or phrase. This translation is generally accomplished by performing a variety of linguistic processing functions that match the phonemic sequences against entries in the system lexicon. The presently preferred linguistic functions include a phonetic dictionary look-up, a context checking function and database, and a basic grammar checking function.” (Hansen, 10:2-11.)

196. The phonetic dictionary includes rules of the English language. (Hansen, 24:39-44 (the phonetic dictionary “contains the English spelling of a word along with its corresponding phonetic representation”), 24:50-52 (the phonetic dictionary “may include a database that contains a more comprehensive Webster-like collection of words”).) The phonetic dictionary maps phonemes to syllables. (Hansen, 10:2-11.)

197. Hansen thus discloses an “*acoustic grammar*” because it includes a grammar of phototactic rules of the English language that maps phonemes to syllables under the Amazon-VBA-Litigation construction and a collection of the phonemes linked together to form syllables, which are linked together to form the words of the language under the Amazon-VBA-IPR interpretation.

198. An exemplary process used by the Linguistic Processor is shown in Figure 10, reproduced below.

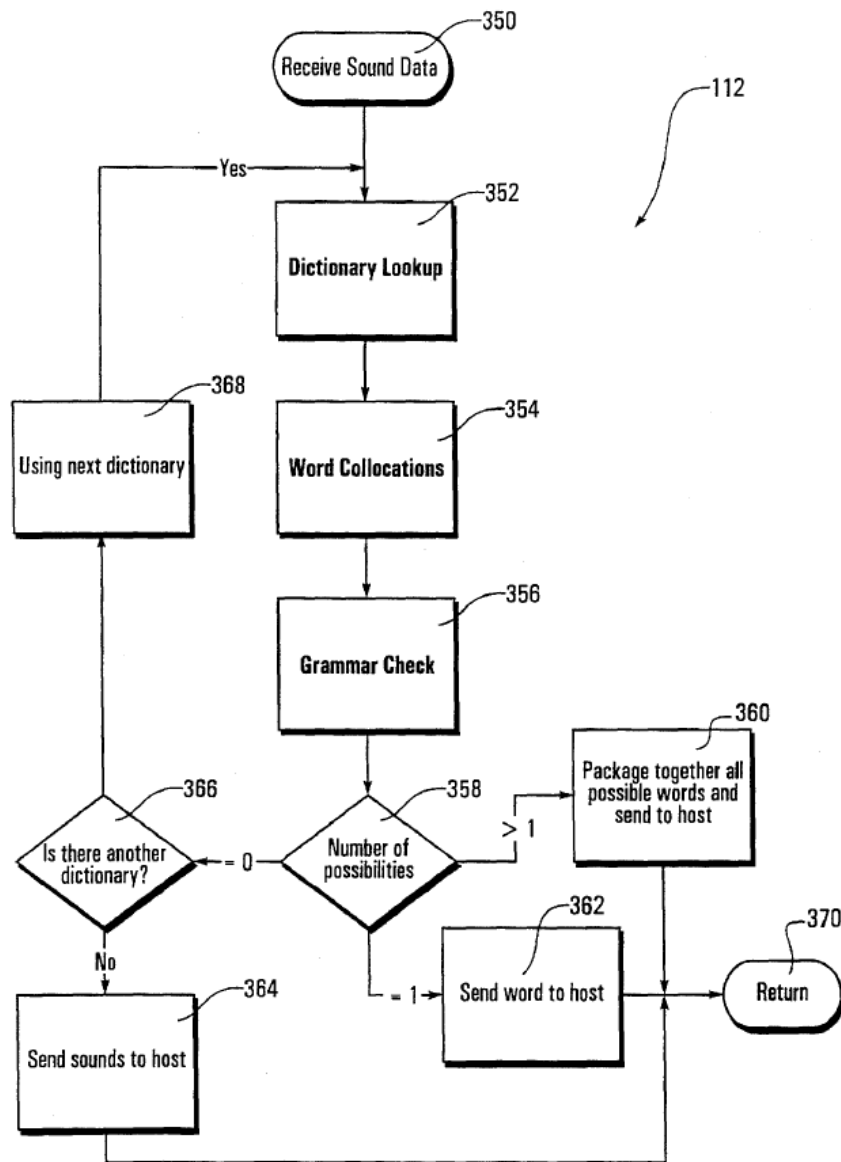


FIG. 10

### Hansen, Figure 10

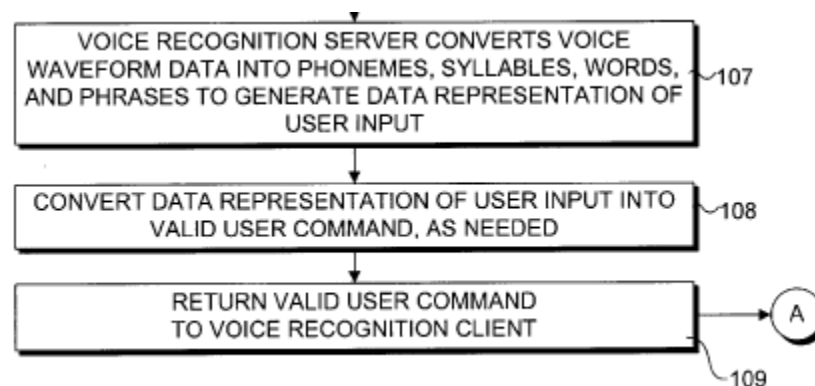
199. "Basically, Dictionary Lookup scans the appropriate dictionary to determine if the incoming sequence of sounds ... form a complete word, or the beginnings of a possible word. To do so, the sounds are placed into paths or

“sequences” to help detect, by way of the phonetic dictionary, the beginning or end of possible words. Thus, as each phoneme sound is received, it is added to the end of a[sic] all non-completed ”sequences.“ Each sequence is compared to the contents of the dictionary to determine if it leads to a possible word.” (Hansen, 24:61-25:4)

The phonetic dictionaries disclosed in Hansen are each an “*acoustic grammar*” e.g., a grammar of phonotactic rules of the English language that maps phonemes to syllables. Phonetic dictionaries that mapped phonemes to syllables were well known in the art. U.S. Patent Publication No. 2007/0067174 (“Verma”; EX-1016) discloses phonemes of speech utterances being mapped to the syllables by using a syllabic mapping database. “The syllabic mapping database maps groups of phones or phonemes to syllables, so that the phones or phonemes of the speech utterance that have been identified can be grouped together into syllables. In this way, by first segmenting the phones or phonemes of a speech utterance and then grouping sequences of the phones or phonemes into syllables, the syllables of the speech utterance are identified.” (Verma, [0025].)

200. Jones ’330, which in discussing functionality in commercially available “speech recognition operations” such as the “Nuance 7.0 speech server” “provided by Nuance Corporation, Menlo Park, Calif.” states that “[t]he phonemes corresponding to the input waveform data are then combined into syllables, words, and/or phrases, as appropriate, based on additional speech recognition data

accessible to voice recognition server 44 (e.g., vocabulary, dictionary, lexicons, etc.) and proprietary algorithms.” (Jones ’330, 7:57-65; 8:56-61) “As individual phonemes are identified, they are chained together to form syllables or single-syllable words. For multi-syllable words, the syllables will then be chained together. When applicable, words are combined into phrases.” (Jones ’330, 8:27-31.) “For example, when a person says the word ”input,“ a speech recognition system will first determine that the word comprises phonemes corresponding to sounds made when pronouncing ”ih,“ ”nnnn“, ”ppp“, ”uhhh,“ and ”ttt“, which may be chained together to form the syllables ”in“ and ”put.“ The corresponding acoustic waveform for ”input“ will include portions corresponding to each phoneme, with little or no blank data in between each phoneme. In contrast, if the user says ”in put“ with a delay between the words, the acoustic waveform data will include a blank portion corresponding to the delay.” (Jones ’330, 8:35-45.) These well-known features are likewise depicted in the Figures of Jones ’330:



**Jones ’330 Figure 5A (excerpt)**

201. The combination of Taschereau and Hansen thus discloses an “*mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar.*”

## 5. Generating Preliminary Interpretation of Utterance [1D] / [27D]

[1D] generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;

[27D] generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;

202. The combination of Taschereau and Hansen discloses “*generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases*” and a speech recognition engine configured to “*generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*”

These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

203. Specifically, Hansen discloses “*generat[ing] a preliminary interpretation for the natural language utterance from the one or more syllables*” during at least three different stages: (i) Dictionary Lookup, (ii) Word Collocations and (iii) Grammar Check. (Hansen, Fig. 10.)

204. “[T]he Dictionary Lookup function utilizes a phonetic-English dictionary that contains the English spelling of a word along with its corresponding phonetic representation. The dictionary can thus be used to identify the English word that corresponds to a particular stream of phonemes.” (Hansen, 24:39-44.) “The phonetic dictionary can be logically separated into several separate dictionaries. For instance, in the preferred embodiment, the first dictionary contains a database of the most commonly used English words. Another dictionary may include a database that contains a more comprehensive Webster-like collection of words. Other dictionaries may be comprised of more specialized words, and may vary depending on the particular application. For instance, there may be a user defined dictionary, a medical dictionary, a legal dictionary, and so on.” (Hansen, 24:46-56.) “Dictionary Lookup scans the appropriate dictionary to determine if the incoming sequence of sounds ... form a complete word, or the beginnings of a possible word.” (Hansen, 24:62-65; Fig. 10A.)

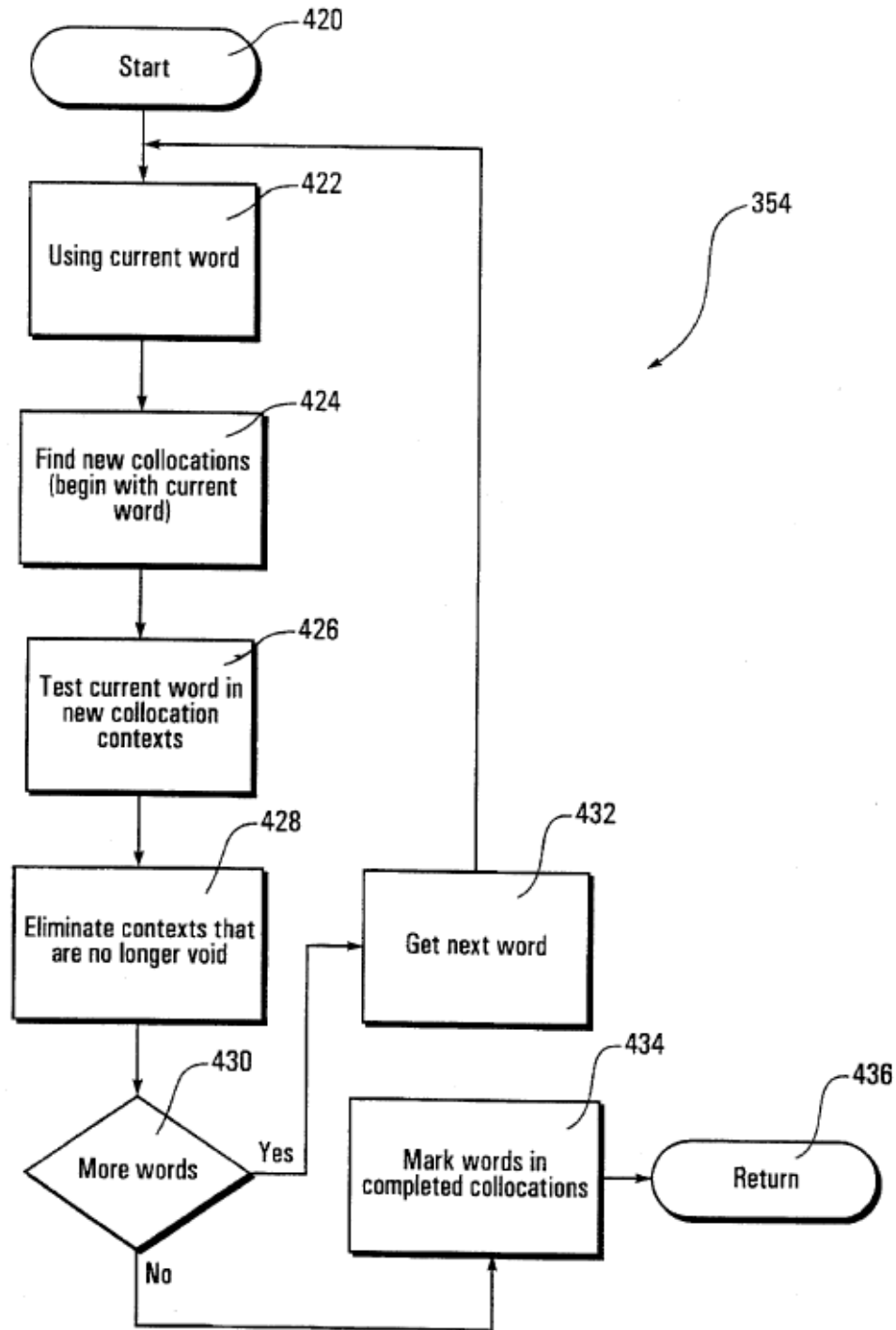
205. Hansen likewise teaches that phonemes are mapped to syllables and from syllables to words. “[A]s each phoneme sound is received, it is added to the

end of a[sic] all non-completed ‘sequences.’” (Hansen, 25:1-2) Hansen equates “sequences” with “syllables.” (Hansen, 25:17-19 (“An ‘active’ sequence is a sequence that could potentially form a word with the addition of a new sound or sounds.”), 25:24-28 (“At program step 388, the host processor 54 checks, by scanning the current [phonetic] dictionary contents, whether the current sequence either forms a word, or whether it could potentially form a word by appending another sound(s) to it.”).) Indeed, in my opinion, a POSITA could not read the reference to “sequences” as referring to anything other than “syllables” or partial words. Additionally, mapping words from syllables was well known in the art as Jones ’330 disclosed “[f]or multi-syllable words, the syllables will then be chained together.” (Jones ’330, 8:28-30.)

206. The phonetic dictionary is used to generate a preliminary interpretation of the utterance. (Hansen, 25:1-7 (“Each sequence is compared to the contents of the dictionary to determine if it leads to a possible word. When a valid word (or set of possible words) is identified, it is passed to the next functional block within the Linguistic Processor portion of the program for further analysis.”).) This block (354) is called “word collocations.” (Hansen, Fig. 10.)

207. “[T]he Word Collocations function monitors the word possibilities that have been identified by the Dictionary Lookup function to see if they form a

‘common’ word collocation.” (Hansen, 26:20-23.) The word collocations analysis process is shown in Figure 10B, reproduced below.



Hansen, Figure 10B

208. “Beginning at program step 420, a set of word possibilities are received. Beginning with one of those words at step 422, the host sound processor 54 next proceeds to program step 424 where it obtains any collocation(s) that have been formed by preceding words. The existence of such collocations would be determined by continuously comparing words and phrases to the collocation dictionary contents. If such a collocation or collocations exist, then the current word possibility is tested to see if it fits within the collocation context. At step 428, those collocations which no longer apply are discarded.” (Hansen, 26:32-42.)

209. “Another optional function that can be used to resolve remaining word possibilities is the ‘Grammar Check’ function, shown at block 356 in FIG. 10. This function evaluates a word possibility by applying certain grammatical rules, and then determining whether the word complies with those rules. Words that do not grammatically fit can be eliminated as possibilities, or assigned lower confidence levels.” (Hansen, 26:53-60.)

210. The Dictionary Lookup, Word Collocations and Grammar Check steps includes a process for “*generat[ing] a preliminary interpretation for the natural language utterance from the one or more syllables.*”

211. Hansen also discloses that the “*the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*” Here, the relevant “recognized words” are the preliminary interpretations generated

by the Dictionary Lookup, Word Collocations and/ or Grammar Check. As noted in Hansen:

In addition to extracting phoneme sound types from the incoming audio speech signal, the sound recognition means implements various linguistic processing techniques to translate the phoneme string into a corresponding **word** or phrase. This can be done for essentially any language that is made up of phoneme sound types.

(Hansen, 3:7-12.)

212. The combination of Taschereau and Hansen discloses “*generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*”

**6. Interpreting the Utterance, Including Establishing Context [1E] / [27E]**

[1E] interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

[27E] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

213. The combination of Taschereau and Hansen discloses “*interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance*” and “*a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

214. The ’176 patent provides describes what it considers “context” when “*interpreting the recognized words or phrases include[ing] establishing a context for the natural language utterance.*” The ’176 patent states that “as described in U.S. patent application Ser. No. 11/200,164, an environmental model may be accessed to determine user location, user activity, track user actions, and/or other environmental information to invoke context, domain knowledge, preferences, and/or other cognitive qualities to enhance the interpretation of questions and/or commands.” (’176 patent, 4:64-5:3.)

215. The ’176 patent provides an example of the system using environmental information, like location, as context for interpreting the user’s natural language

utterance: “Based on a context of the requested task (e.g., ”Navigation,“ ”Seattle,“ etc.), a voice search engine may retrieve an address of the Space Needle and a navigation application may calculate the route. Furthermore, user profile information may indicate that the user is visiting Seattle from out-of-town (e.g., the profile may indicate that the user’s home is Sacramento), and therefore, an advertisement for popular points-of-interest in Seattle may be selected.” (’176 patent, 8:55-63.)

216. Similar to the ’176 patent, Taschereau discloses interpreting the recognized words at a “*conversational language processor*.”

217. Taschereau discloses an “adaptive” system that “can take advantage of linguistics ... and can extract meaning from the utterance.” (Taschereau, 22:23; 23:6-7.) Unlike “[p]rior art ASR systems [that] have been concentrating on ‘what was said,’” the Taschereau system focuses on determining “what was meant.” (Taschereau, 23:7-9.)

218. Taschereau achieves this goal by interpreting the recognized words and phrases with a multi-pass process using a “dynamically generated ‘latent grammar.’” (Taschereau, 14:7-8.) The “stored utterance” (preliminary interpretation from the word list) is “re-input against the latent recognition grammar.” (Taschereau, 15:22-24.) The utterance can be “run through the latent grammar (more than once if necessary) and an answer is provided.” (Taschereau 17:11-12; *see also* 15:22-24.)

“Grammars can further be broken down into very specific classes, for example all of the pizza restaurants in a given locality, or all of the hotels.” (Taschereau, 19:1-4.) Certain words in the “latent recognition grammar” “may trigger geographic referencing (such as the term ‘on’) which can be used by the system for accuracy.” (Taschereau, 21:20-24.) Because the multi-pass latent recognition process recognizes words and phrases, it is a “*conversational language processor*.”

219. Taschereau also discloses “*wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance*.” Like the ’176 patent, Taschereau can use environmental information, like location, as context to interpret the recognized words of the natural language utterance: “The system and method described above can also serve to direct services to users or direct users to services. For example when a user requests the phone number of a taxi company, it is likely that user is actually trying to have a taxi sent to a particular location. The ASR system can be used with geographic recognition systems.” (Taschereau, 23:19-22.) For example, the system may “ask the user if they are looking for a service, e.g. a taxi, or the nearest hotel, and if so, they can be asked to give their location. Then after determining the location of the user they can be directed to the nearest hotel, or the closest taxi can be directed to them. This feature can be used with a number of services, including restaurants, pizza, laundromats, etc.” (Taschereau, 23:24-24:2.)

220. Thus, in my opinion [a POSITA would have understood that] Taschereau's multi pass process is a conversational language processor interpreting the recognized words and phrases using grammars broken down into very specific classes such as pizza restaurants or hotels in a given locality and a dynamically generated latent grammar. Also, [A POSITA would have also understood that] the conversational language processor establishes a context for the natural language utterance, such as a taxi request, or the nearest hotel by directing services to users or viceversa.

221. Taschereau further discloses other ways the system can use location information: "The geographic referencing can also be used to provide answers when the user gives incorrect information. For example, if the user asks for a listing that doesn't exist in a particular location, the system can look in neighbouring areas (for example a suburb) to determine if the appropriate listing is actually there. Also areas that have very similar sounds may be checked. For example if a reference can't be located in the town named 'Oshawa,' the ASR system, time permitting can, then check the location 'Ottawa'." (Taschereau, 24:3-8.)

222. Hansen likewise discloses "*interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine.*"

223. “As words are extracted from the incoming speech signal by the Linguistic Processor, they are immediately passed to the next function in the overall program method referred to as the ‘Command Processor.’” (Hansen, 27:34-37.) “In the preferred embodiment, the Command Processor is a series of program steps that are executed by a Host Computer 22.” (Hansen, 27:38-40.) “The Command Processor then receives each word, and determines the manner by which it should be used on the host computer 22. For example, a spoken word may be input as text directly into an application, such as a wordprocessor document. Conversely, the spoken word may be passed as a command to the operating system or application.” (Hansen, 27:44-49.)

224. Hansen explains the process for identifying the nature of a word thusly:

[P]rogram step 450 causes the host computer 22 to receive a word created by the Linguistic Processor portion of the algorithm. The host computer 22 then determines, at step 452, whether the word received is an operating system command. This is done by comparing the word to the contents of a definition file database, which defines all words that constitute operating system commands. If such a command word is received, it is passed directly to the host computer 22 operating system, as is shown at program step 454.

If the incoming word does not constitute an operating system command, step 456 is executed, where it is determined if the word is instead an application command, as for instance a command to a wordprocessor or spreadsheet. Again, this determination is made by comparing the word to another definition file database, which defines all words that constitute an application

command. If the word is an application command word, then it is passed directly, at step 458, to the intended application.

If the incoming word is neither a operating system command, or an application command, then program step 460 is executed, where it is determined whether the Command Processor is still in a “command mode.” If so, the word is discarded at step 464, and essentially ignored. However, if the Command Processor is not in a command mode, then the word will be sent directly to the current application as text.

(Hansen, 27:52-28:11.)

225. The “Command Processor” and associated host computer together constitute the claimed “*conversational language processor*” because they are coupled to the speech recognition engine and interpret the recognized words or phrases.

226. Hansen also discloses the step “*wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*” As noted above, the “Command Processor” and associated host computer interpret the words and phrases generated by the (i) Dictionary Lookup, (ii) Word Collocations and/or (iii) Grammar Check. The “Command Processor” and associated host computer then “determine[] the manner by which [the supplied word] should be used on the host computer 22.” (Hansen, 27:44-49.) As noted above, the “Command Processor” and associated host computer determine whether the word or phrase is an “operating system command,” “application command” or

“neither a[sic] operating system command, or an application command.” (*Supra* §V.A.2.) Each is a “*context for the natural language utterance.*”

227. The combination of Taschereau and Hansen thus discloses “*interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*”

#### **7. Selecting An Advertisement [1F] / [27F]**

[1F] selecting an advertisement in the context established for the natural language utterance; and

[27F] select an advertisement in the context established for the natural language utterance; and

228. The combination of Taschereau and Hansen discloses “*selecting an advertisement in the context established for the natural language utterance*” and a conversational language processor configured to “*select an advertisement in the context established for the natural language utterance.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

229. Taschereau discloses using the established context to select a relevant advertisement for the natural language utterance. For example, Taschereau describes using information about the user and information about the user’s request to select the advertisement:

The making of a request for a business also provides an opportunity to target an advertisement. For example when a request is made for a restaurant in a certain geographic area, a competitor could present an advertisement with an inducement (e.g. a coupon or the like) in an attempt to lure that customer to a different establishment. By using the information available about the user and the listing the user is looking for, very precise advertisements can be presented to the user.

(Taschereau, 25:5-11.)

As discussed above, the “information about the user” can be environmental information about the user, like the user’s location. Therefore, Taschereau discloses “*selecting an advertisement*” (like a local restaurant) “*in the context established for the natural language utterance*” (the user’s location in the area of the local restaurant).

230. Another example disclosed in Taschereau relates to “geographic recognition systems” used with the ASR system. (Taschereau, 23:21-24.) The Taschereau system “can be modified to ask the user if they are looking for a service, e.g., a taxi, or the nearest hotel, and if so, they can be asked to give their location. Then after determining the location of the user they can be directed to the nearest hotel, or the closest taxi can be directed to them. This feature can be used with a number of services, including restaurants, pizza, laundromats, etc.” (Taschereau, 23:25-24:2.)

231. The combination of Taschereau and Hansen thus discloses “*selecting an advertisement in the context established for the natural language utterance.*”

#### 8. Presenting the Selected Advertisement [1G] / [27G]

[1G] **presenting the selected advertisement via an output device** coupled to the conversational language processor.

[27G] **present the selected advertisement via an output device.**

232. The combination of Taschereau and Hansen discloses “*presenting the selected advertisement via an output device coupled to the conversational language processor*” and a conversational language processor configured to “*present the selected advertisement via an output device.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

233. Taschereau claims this limitation in claim 9:

A method of providing directory assistance to a user comprising:

(a) receiving an utterance from a user;

(b) determining a listing in response to said utterance;

(c) **providing an advertisement to said user** before providing said listing to said user;

wherein said user is not charged an additional fee for the directory assistance.

(Taschereau, claim 9.)

234. Taschereau further describes when and how advertisements are presented to the user:

The making of a request for a business also provides an opportunity to target an advertisement. For example when a request is made for a restaurant in a certain geographic area, **a competitor could present an advertisement** with an inducement (e.g. a coupon or the like) in an attempt to lure that customer to a different establishment. The user will also be providing information about themselves based on the area from which they are calling and the call display information. **By using the information available about the user and the listing the user is looking for, very precise advertisements can be presented to the user.**

(Taschereau, 25:5-11.)

235. Taschereau further describes that “[c]ertain words will also imply geographical relationships. For example the words ‘in’, ‘around’, ‘near’ and ‘on’ usually imply such a relationship. In these cases several advantages may be accomplished. **For example, if the directory assistance service is advertiser supported a very targeted advertisement may be selected to play to the user.**”

(Taschereau, 14:16-19.)

236. A POSITA would understand that the targeted advertisement is played to the user or provided to the user “*via an output device coupled to the conversational language processor*” (e.g., audibly through the telephone).

237. The combination of Taschereau and Hansen discloses “*presenting the selected advertisement via an output device coupled to the conversational language processor*” and a conversational language processor configured to “*present the selected advertisement via an output device.*”

### C. Dependent Claims 2 and 28

2. The method of claim 1, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

28. The system of claim 27, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

238. Dependent claims 2 and 28 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 2 and 28 add the following limitation: “*wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.*”

239. Taschereau discloses “*wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.*” Specifically, Taschereau discloses a “*conversational language processor [that] selects the advertisement based on*” at least both (i) “*information related to one or more of the recognized words or phrases*” and (ii) “*an environmental model derived from environmental conditions or surroundings associated with the specific user.*”

240. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.” (*Supra* §IV.D.) I have applied that construction in my analysis of these claims.

241. As discussed above, Taschereau discloses a system that interprets the user’s natural language request and uses the interpretation to select and present an advertisement to the user. (*See supra* §§V.B.7-8.) Taschereau provides at least one example in which “a request is made for a restaurant in a certain geographic area, a

competitor could present an advertisement with an inducement (e.g. a coupon or the like) in an attempt to lure that customer to a different establishment.” (Taschereau, 25:5-9.) In this example, the system is using “*one or more of the recognized words*” (the name of the restaurant the user requested) to select an advertisement (the advertisement for the competing restaurant).

242. Also as discussed above, Taschereau discloses the use of geographic information to select and present advertisements to users. (*See supra* §§ V.B.7-8.) The Taschereau system “can be modified to ask the user if they are looking for a service, e.g., a taxi, or the nearest hotel, and if so, they can be asked to give their location. Then **after determining the location of the user they can be directed to the nearest hotel, or the closest taxi can be directed to them.** This feature can be used with a number of services, including restaurants, pizza, laundromats, etc.” (Taschereau, 23:25-24:2.)

243. The ’176 patent describes an “*environmental model*” as follows: “For example, as described in U.S. patent application Ser. No. 11/200,164, an environmental model may be accessed to determine **user location**, user activity, track user actions, and/or other environmental information to invoke context, domain knowledge, preferences, and/or other cognitive qualities to enhance the interpretation of questions and/or commands.” (’176 patent, 4:64-5:3.) In Taschereau, as described above, the system determines a “user location” to enhance

the interpretation of the user's utterances. A POSITA would understand that the user location information being accessed in the Taschereau is part of an "*environmental model*" as described in the '176 patent because it is an approximation or representation of an aspect of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system. Taschereau therefore discloses a "*model*" under the Amazon-VBA-Litigation construction.

244. Moreover, the '176 patent admits that the use of an environmental model "to enhance the interpretation of questions and/or commands" was already known in the art and not a novel feature of the alleged invention. ('176 patent, 4:64-5:3.)

245. The combination of Taschereau and Hansen discloses "*wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific use.*"

#### D. Dependent Claims 9 and 35

9. The method of claim 1, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

35. The system of claim 27, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

246. Dependent claims 9 and 35 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 9 and 35 add the following limitation: “*wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.*”

247. Taschereau discloses that the system can resolve the user’s request when the user provides ambiguous information: “Also areas that have very similar sounds may be checked. For example if a reference can’t be located in the town named ‘Oshawa’, the ASR system, time permitting can, then check the location ‘Ottawa’.” (Taschereau, 24:6-8.) When Taschereau describes “areas that have very similar sounds,” I understand Taschereau describing ambiguous information in the natural language utterance. Taschereau describes providing advertisements to user’s in response to the user making a request for a business in a particular area.

(Taschereau, 25:5-11.) In the example above, the conversational language processor in Taschereau would select an advertisement for “Ottawa” after determining that “Oshawa” could not be located because it is incomplete or at least ambiguous. When the system selects the advertisement related to “Ottawa” the system will have *“resolve[d] the request in response to determining that the natural language utterance includes incomplete or ambiguous information.”*

248. Taschereau discloses that the disclosed system can resolve the user’s request when the user provides incorrect information: “The geographic referencing can also be used to provide answers when the user gives incorrect information. For example, if the user asks for a listing that doesn’t exist in a particular location, the system can look in neighbouring areas (for example a suburb) to determine if the appropriate listing is actually there.” (Taschereau 24:3-6.) In my opinion, Taschereau’s ability to resolve incorrect requests is resolving *“incomplete or ambiguous information.”* In both instances, the conversational language processor must determine that the natural language utterance does not contain precise information related to the request and resolve the natural language utterance to select and provide an advertisement to the user based on the request.

249. The combination of Taschereau and Hansen discloses *“wherein the conversational language processor selects the advertisement to resolve the request*

*in response to determining that the natural language utterance includes incomplete or ambiguous information.”*

**E. Dependent Claims 10 and 36**

10. The method of claim 1, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

36. The system of claim 27 wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

250. Dependent claims 10 and 36 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 10 and 36 add the following limitation: “*wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.*”

251. As discussed above, Taschereau uses a “*speech recognition engine*” to “*recognize one or more words or phrases in a natural language utterance.*” (See *supra* §V.B.3.) Taschereau discloses one way to do so is to use a “Word List” that

is intended to be a collection of “*one or more words or phrases in the natural language utterance.*” Taschereau further discloses a “grammar generation process” that is part of its “ASR system” and is used to “generate a single list of terms” that can be used in the speech recognition system. (Taschereau, 7:24-25; 10:21-24.) The ASR system in Taschereau discloses that words and phrases in the natural language utterance are recognized in dictionary and phrase tables as the Word List. (*See supra* §§V.A.1, V.B.3.) The “Word List ... is assembled into an ASR grammar.” (Taschereau, 12:20-21, Abstract.)

252. During speech recognition, the utterance is applied against the grammar and the “n-best” matches for a given utterance may be returned.” (Taschereau, 13:21-28.) All entries of the n-best words for a given utterance are based on “[a] list of grammar results and associated return result frequency and confidence scores[.]” (Taschereau, 13:28-29.) The Word List is scanned and “all entries containing any of the n-best words ... are placed in a dynamically generated ‘latent grammar.’” (Taschereau, 14:8-9.) Additionally, Taschereau discloses that “[i]t may be herlpful [sic] to add other words to the grammar, such as those associated with frequently request[ed] [sic] listings.” (Taschereau, 14:11-12.) “Likewise, the presence of certain words will imply that others should be included such as when the word “mall” is recognized, the term “shopping center” should also be added.” (Taschereau, 14:13-14.) An exemplary list of “n-best” words and confidence levels

generated in response to a request for “Kearney Funeral Home” is shown in Figure 12, below.

Session ID: 228961151  
Country Input: ca  
Country Confidence Level:0.76  
City Input: geopl\_c\_common\_12459\_0000000149  
City Confidence Level:0.77  
Type Input: business  
Type Confidence Level:0.75  
Name Confidence Level: 0.6  
Number of Name NBest: 5  
Name Input 1: chair  
Name Input 2: nishio  
Name Input 3: orear  
Name Input 4: palm  
Name Input 5:  
Name Input 6:  
NBest Results 1: chair,chair,chair,chair,chair  
NBest Results 2: nishio,nishio,nishio,nishio,nishio  
NBest Results 3: orear,orear,arrow,aero,arrow  
NBest Results 4: palm,pomme,home,home,pomme  
NBest Results 5:  
NBest Results 6:

### **Taschereau, Figure 12**

253. The Word List is scanned and “all entries containing any of the n-best words ... are placed in a dynamically generated ‘latent grammar.’” (Taschereau, 14:8-9.) “The utterance is then run through the latent grammar (more than once if necessary) and an answer is provided.” (Taschereau, 17:1-12.) Taschereau thus discloses a “*speech recognition engine [that] recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables*” where the “latent grammar” is the “*plurality of entries*

*in one or more dictionary and phrase tables.*” (Taschereau, 14:26-15:2 (giving exemplary phrase tables).)

254. The “latent grammar” is also “*dynamically updated based on a history of a current dialog and one or more prior dialogs.*” As noted above, “n-best” words (which comprise the “latent grammar”) are determined based, in part, on “associated return result frequency and confidence scores” which are created based on prior dialogs. (Taschereau, 13:28-29.) “Calculating the result occurrence frequency and obtaining the confidence score can be applied in a number of ways to effectively determine the relevance of items in the result set.” (Taschereau, 13:29-14:1.) “The latent grammars may be dynamically or previously generated.” (Taschereau, 7:3.)

255. “The grammar is dynamically generated knowing that the ASR system will be used again to perform one or more latent, and optionally concurrent, recognitions, each latent recognition evaluating the terms from a previous recognition process.” (Taschereau, 6:16-20.) The grammar can also be changed manually: “[i]t may be herlpful [sic] to add other words to the grammar, such as those associated with frequently request[ed] [sic] listings.” (Taschereau, 14:11-12.) A POSITA would understand that dynamically generating the grammar includes updating a plurality of entries in a dictionary or phrase table based on the current dialog (n-best words for a given utterance) and one or more prior dialogs which contained a frequently requested listing. The combination of Taschereau and Hansen

thus discloses “*wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.*”

**F. Dependent Claims 11 and 37**

11. The method of claim 10, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

37. The system of claim 36, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

256. Dependent claims 11 and 37 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 11 and 37 add the following limitation: “*wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.*”

257. As discussed above, Taschereau discloses that the entries in the Word List (the dictionary) are “*dynamically updated based on a history of a current dialog and one or more prior dialogs.*” (See *supra* §V.E.) As discussed above, the dynamic

updating described in Taschereau is done based on the frequency of usage of terms. (See *supra* §V.A.1.) A POSITA would understand that Taschereau's dynamic updating based on the frequency of the usage of terms is dynamically updating "based on ... prior probabilities derived from the history of the current dialog and the prior dialogs" because the system is using probabilities (frequency of use of a term) to dynamically update the Word List.

258. The combination of Taschereau and Hansen discloses "*wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.*"

## **VI. GROUND 2: THE COMBINATION OF TASCHEREAU, HANSEN AND FRANCO RENDERS OBVIOUS CLAIMS 12-18, 25-26, 38-44, AND 51-52**

### **A. Overview of the Combination**

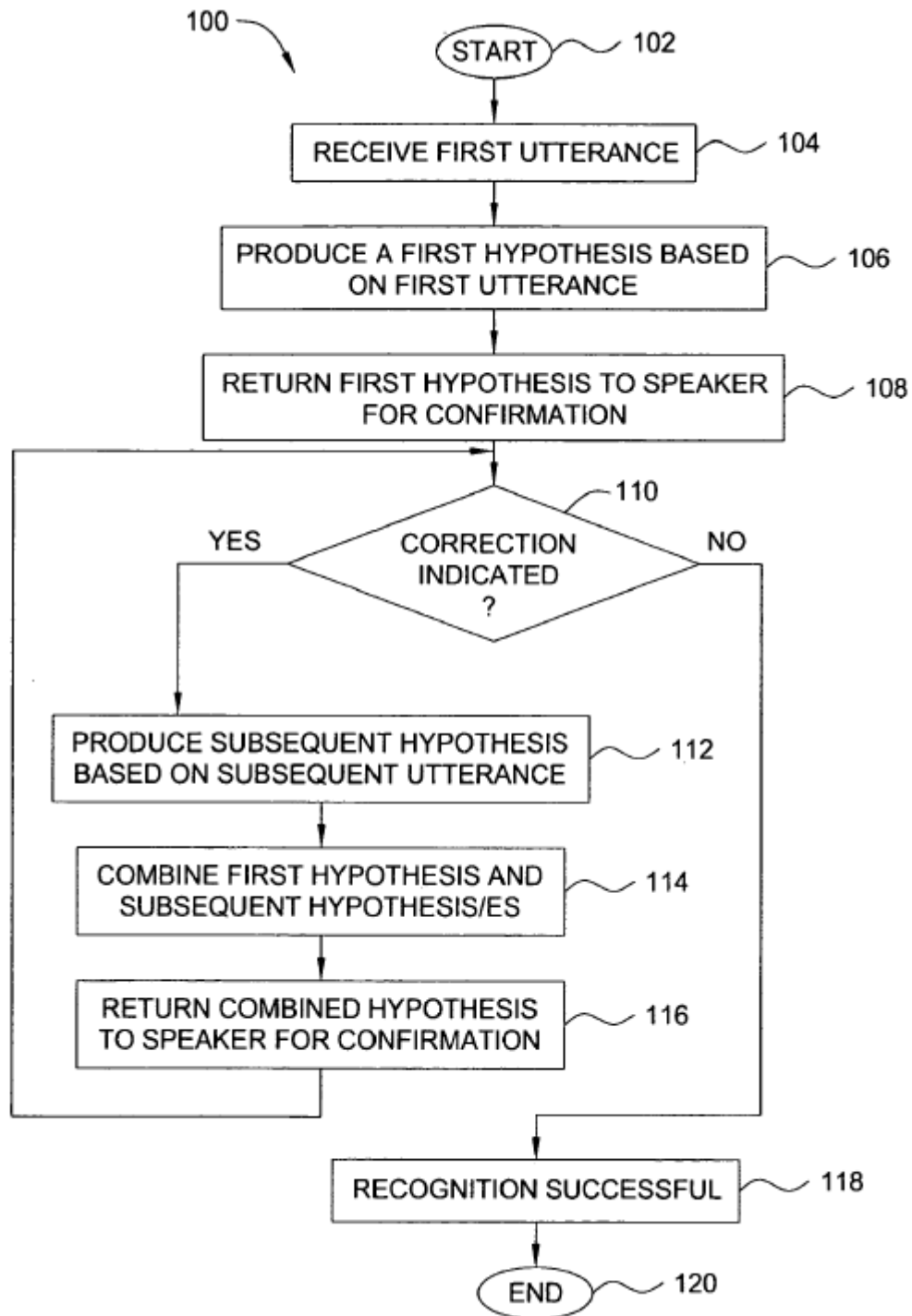
259. The combination of Taschereau, Hansen and U.S. Patent Application Publication No. US 2010/0125458 ("Franco"; EX-1007) discloses every limitation of claims 12-18, 25-26, 38-44, and 51-52. Taschereau, Hansen and Franco are analogous art to the '176 patent because all three references are directed to speech recognition systems and software.

260. I have described both Taschereau and Hansen above, and I discuss Franco below.

## **1. Franco**

261. Franco “relates generally to speech recognition applications and relates more particularly to correcting errors in speech recognition systems.” (Franco, [0002].)

262. Franco includes multiple embodiments for determining that a word or phrase has been incorrectly interpreted. In one embodiment (shown in Figure 1 below), an error is determined based on feedback from a user. (Franco, [0013]-[0015].) The user provides a “‘correction’ utterance” and the system generates a “combined hypothesis” on what is being requested. (Franco, [0017].)



**Franco, Figure 1**

263. Another embodiment relates to a scenario where a user offers an utterance, the utterance is incorrectly recognized, and the user repeats the same

utterance without being prompted to do so by the system. (Franco, [0018].) In such scenario, the method combines hypothesized interpretations of the first and second (repeated) utterance to maximize the likelihood of correct interpretation. (Franco, [0020] (“The method 100 thereby achieves improved error correction for speech recognition applications. By assuming that a subsequent speaker utterance will comprise all or part of a previously spoken (**and incorrectly recognized**) utterance, the acoustic evidence from the previous and subsequent utterances can be combined in an efficient manner to produce a more accurate corrected recognition.”).) For example, a user may state “[t]he number is 1234. I repeat, the number is 1234”. (Franco, [0018].) “In this case, the method 100 would treat ‘The number is 1234’ as the first utterance and ‘I repeat, the number is 1234’ as the subsequent utterance, and combine respective first and subsequent hypotheses [] to produce the combined hypothesis” for recognizing the speech. (Franco, [0020].)

264. Franco thus discloses a method that “includes receiving a first utterance from the user, receiving a subsequent utterance from the user, and combining acoustic evidence from the first utterance with acoustic evidence from the subsequent utterance in order to recognize the first utterance.” (Franco, [0006].) “It is assumed that, if the first utterance has been incorrectly recognized on a first attempt, the user will repeat the first utterance (or at least the incorrectly recognized portion of the first utterance) in the subsequent utterance.” (Franco, [0006].)

## **2. Motivation to Combine**

265. A POSITA would have been motivated to integrate the speech recognition error correction features disclosed in Franco with the speech recognition systems disclosed in Taschereau and Hansen for a number of reasons.

266. First, Taschereau recognizes the motivation to combine: “[a] significant limitation with ASR systems in the prior art is that as a grammar’s size increases, its accuracy diminishes. This occurs because as the number of possible phonetic matches for an utterance increase, the probability for error also increases as the differences between the various possible matches is smaller, (i.e. each possible match become less distinct).” (Taschereau, 2:1-4.) A POSITA would understand Taschereau expressing a motivation to combine the system disclosed in Taschereau with additional systems, like Franco, that specialize in correcting errors in speech recognition systems.

267. Second, Franco also recognizes the motivation, stating that there is a “need in the art for a method and apparatus for error correction in speech recognition applications.” (Franco, [0005].) A POSITA would understand that Franco is suggesting that the method of error correction disclosed in Franco was broadly applicable and could be implemented and combined with speech recognition applications, like Taschereau and Hansen.

268. Third, a POSITA would have been motivated to modify the recursive process in Taschereau (which includes “multiple simultaneous passes of the same utterance” (Taschereau, 19:20-20:7)) with identified “predetermined events” like a user repeating the same utterance in order to reduce processing time.

269. Fourth, the results of the combination would have been predictable to a POSITA and the POSITA would have had a reasonable expectation of success because Franco discloses that its error recognition module can be “implemented as a physical device or subsystem that is coupled to a processor through a communication channel.” (Franco, [0027].) Taschereau is implemented through the telephone system, an example of a communication channel.

#### **B. Independent Claims 14 and 40**

270. Independent claims 14 and 40 include substantially overlapping processing limitations ([14A]-[14F]/[40A]-[40F]), as shown in the table below. Because of the overlap, I address these limitations together. Moreover, all but one of these limitations overlap with limitations addressed in independent claims 1 and 14 so have already been addressed above. In this table, I also provide a citation to the section discussing each limitation.

| <b>Claim 1</b>                                                                                                                                                                                                                                               | <b>Claim 27</b>                                                                                                                                                                                                                                                                                                         | <b>Section</b> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| [14P] A method for selecting and presenting advertisements in response to processing natural language utterances, comprising:                                                                                                                                | [40P] A system for selecting and presenting advertisements in response to processing natural language utterances, comprising:                                                                                                                                                                                           | §VI.B.1        |
| [14A] receiving a natural language utterance containing at least one request at an input device;                                                                                                                                                             | [40A] an input device that receives a natural language utterance containing at least one request at an input device;                                                                                                                                                                                                    | §VI.B.2        |
| [14B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device;                                                                                                                 | [40B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance;                                                                                                                                         | §VI.B.3        |
| [14C] interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | [40C] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | §VI.B.4        |
| [14D] selecting an advertisement in the context established for the natural language utterance;                                                                                                                                                              | [40D] select an advertisement in the context established for the natural language utterance; and                                                                                                                                                                                                                        | §VI.B.5        |

| Claim 1                                                                                                                                                                                                                                                                                                             | Claim 27                                                                                                                                                                                                                                                                                                              | Section |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| [14E] presenting the selected advertisement via an output device coupled to the conversational language processor; and                                                                                                                                                                                              | [40E] present the selected advertisement via an output device; and                                                                                                                                                                                                                                                    | §VI.B.6 |
| [14F] determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event. | [40F] an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event. | §VI.B.7 |

## 1. Preamble [14P] / [40P]

[14P] A method **for selecting and presenting advertisements in response to processing natural language utterances, comprising:**

[40P] A system **for selecting and presenting advertisements in response to processing natural language utterances, comprising:**

271. The combination of Taschereau and Hansen discloses “[a] *method for selecting and presenting advertisements in response to processing natural language utterances*” as well as a “*system for selecting and presenting advertisements in response to processing natural language utterances,*” as discussed below in §§VI.B.1-7. The preambles for independent claims 14 and 40 are the exact same as

those for claims 1 and 27, respectively. For the same reasons discussed above in §V.B.1 the combination of Taschereau and Hansen discloses these limitations.

## 2. Receiving Natural Language Request [14A] / [40A]

[14A] receiving a natural language utterance containing at least one request at an input device;

[40A] an input device that receives a natural language utterance containing at least one request at an input device;

272. The combination of Taschereau and Hansen discloses “*receiving a natural language utterance containing at least one request at an input device*” as well as a “*an input device that receives a natural language utterance containing at least one request at an input device.*” Claim limitations [14A] and [40A] are the same as those for claim limitations [1A] and [27A], respectively. For the same reasons discussed above in §V.B.2, the combination of Taschereau and Hansen discloses these limitations.

## 3. Recognizing One or More Words in the Utterance [14B] / [40B]

[14B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device;

[40B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance;

273. The combination of Taschereau and Hansen discloses “*recognizing one or more words or phrases in the natural language utterance at a speech recognition*

*engine coupled to the input device” as well as a “a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance.”* Claim limitations [14B] and [40B] are the same as those for claim limitations [1B] and [27B], respectively. For the same reasons discussed above in §V.B.3, the combination of Taschereau and Hansen discloses these limitations.

#### **4. Interpreting the Utterance Using Context [14C] / [40C]**

[14C] **interpreting the recognized words or phrases** at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

[40C] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: **interpret the recognized words or phrases**, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

274. The combination of Taschereau and Hansen discloses “*interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance*” as well as a “*a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases*

*includes establishing a context for the natural language utterance.”* Claim limitations [14C] and [40C] are the same as those for claim limitations [1E] and [27E], respectively. For the same reasons discussed above in §V.B.6, the combination of Taschereau and Hansen discloses these limitations.

## **5. Selecting An Advertisement [14D] / [40D]**

[14D] selecting an advertisement in the context established for the natural language utterance;

[40D] select an advertisement in the context established for the natural language utterance; and

275. The combination of Taschereau and Hansen discloses “*selecting an advertisement in the context established for the natural language utterance*” as well as a conversational language processor configured to “*select an advertisement in the context established for the natural language utterance.*” Claim limitations [14D] and [40D] are the same as those for claim limitations [1F] and [27F], respectively. For the same reasons discussed above in §V.B.7, the combination of Taschereau and Hansen discloses these limitations.

## **6. Presenting the Selected Advertisement [14E] / [40E]**

[14E] presenting the selected advertisement via an output device coupled to the conversational language processor; and

[40E] present the selected advertisement via an output device; and

276. The combination of Taschereau and Hansen discloses “*presenting the selected advertisement via an output device coupled to the conversational language processor*” as well as a conversational language processor configured to “*present the selected advertisement via an output device.*” Claim limitations [14E] and [40E] are the same as those for claim limitations [1G] and [27G], respectively. For the same reasons discussed above in §V.B.8, the combination of Taschereau and Hansen discloses these limitations.

**7. Determining An Incorrect Interpretation Occurred and Reinterpreting [14F] / [40F]**

[14F] determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

[40F] an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

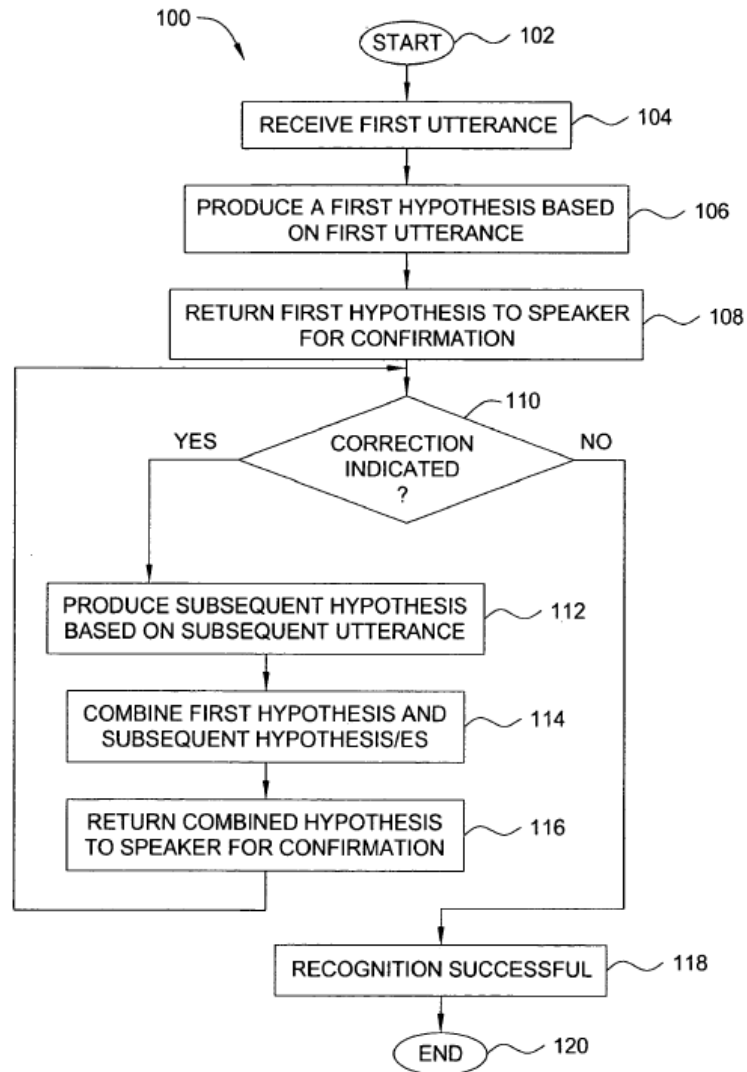
277. The combination of Taschereau, Hansen and Franco discloses “*determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event*” as well as “*an adaptive*

*misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.*” Franco discloses these limitations.

278. Franco discloses an “*adaptive misrecognition engine*” that determines when a “*conversational language processor incorrectly interpret[s] words or phrases*” wherein the “*conversational language processor reinterprets the words or phrases in response to [a] predetermined event.*”

279. Franco “relates generally to speech recognition applications and relates more particularly to correcting errors in speech recognition systems.” (Franco, [0002].) Errors can be found when, for example, “background noise present at the time the utterance is made or speaker enunciation [] affect[s] a speech recognition application’s ability to produce accurate results in a timely fashion.” (Franco, [0004].) As I discussed above, in one Franco embodiment, a user offers an utterance, the utterance is “incorrectly recognized” by the system, and the user repeats the same utterance without being prompted to do so by the system. (Franco, [0018], [0020].) Because the system is able to determine that it “incorrectly recognized” the first utterance, it includes an “*adaptive misrecognition engine.*”

280. In another embodiment also involving successive utterances (shown in Figure 1, below), the system receives a first utterance (104) and provides a hypothesized response to the speaker (108). (Franco, Figure 1.)



**Franco, Figure 1**

281. If the speaker indicates a correction is needed, the system then generates a hypothesis for the second utterance **and a new, combined hypothesis for the first utterance** based on the first and second utterances. (Franco, [0017].)

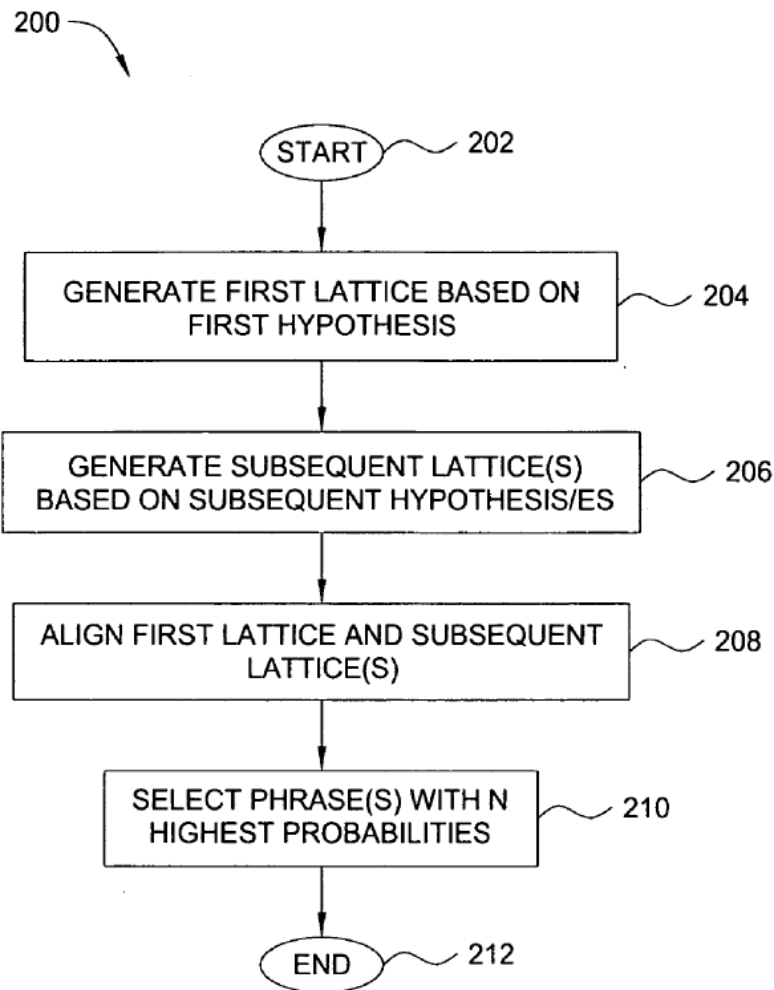
Because in each of these examples the system is able to determine that it “incorrectly recognized” the first utterance, Franco discloses an “*adaptive misrecognition engine*.”

282. Franco also discloses the limitation requiring that incorrect interpretation be made “*in response to an adaptive misrecognition engine detecting a predetermined event*.” Franco discloses error correction in the specific situation where a speaker naturally repeats an utterance (e.g., “The number is 1234. I repeat, the number is 1234”). (Franco, [0018].) Franco’s system assumes that “when speech recognition is unsuccessful on the first attempt, a speaker will naturally repeat all or part of the same utterance in order to correct the recognition error.” (Franco, [0012].) Franco likewise discusses error detection when specific keywords are used like “repeating.” (Franco, [0015] (“In one embodiment, the need for a correction is indicated by a subsequent speaker utterance including a predefined keyword.”).) A speaker repeating all or part of the same utterance or reciting a specific keyword are each a “*predetermined event*” that the adaptive misrecognition engine detects.

283. Franco discloses “*the conversational language processor reinterprets the words or phrases in response to the predetermined event*” in at least two ways. In the embodiment where the speaker repeats all or part of the same utterance “without an intervening prompt from the method 100,” a hypothesis is presented “to the speaker for review/confirmation.” (Franco, [0018]-[0019].) If “further correction

to the combined hypothesis is necessary,” the system will “continue to combine hypotheses in an effort to correct recognition **of the first utterance** until recognition is determined to be successful” as shown in Figure 1 above. (Franco, [0019].)

284. In another embodiment (figure 2, below), Franco discloses combining hypotheses by first generating word lattices corresponding to the hypotheses.



**Franco, Figure 2**

285. After aligning lattices, the system can “select the phrase(s) with the N highest probabilities as the combined hypothesis to be presented to the speaker.”

(Franco, [0026].) Alternatively, “if the phrase with the highest probability is identical to the first hypothesis, which the speaker has indicated as being incorrect, then the method 200 selects the phrase with the second highest probability as an alternative combined hypothesis.” (Franco, [0026].)

286. In both examples, the system reinterprets the misrecognized words from the first utterance in response to the predetermined event (the repetition of the utterance or use of a keyword by the user indicative of a recognition error) by using the first hypothesis in subsequent hypotheses. (Franco, [0020] (“the acoustic evidence from the previous and subsequent utterances can be combined in an efficient manner to produce a more accurate corrected recognition”).)

287. This disclosure is consistent with the ‘176 patent intrinsic record. I understand that, in the specification, the applicants referenced “U.S. patent application Ser. No. 11/200,164, entitled ‘System and Method of Supporting Adaptive Misrecognition in Conversational Speech,’ ... which issued as U.S. Pat. No. 7,620,549” [“Di Cristo”]. (‘176 patent, 4:55-64.)

288. Di Cristo explains that the one way the “adaptive misrecognition engine” works is by running through a list of highest scored interpretations and presenting them serially to the user. (Di Cristo, 33:42-48 (when the “adaptive misrecognition engine determines that the interpretation generated” is incorrect, the system should “select the next best interpretation from the plurality of

interpretations”).) Franco thus discloses a process for reinterpreting misrecognized utterances that mirrors the support for the alleged invention in the ’176 patent.

289. The system in Franco assumes that “when speech recognition is unsuccessful on the first attempt, a speaker will naturally repeat all or part of the same utterance in order to correct the recognition error. Thus, acoustic evidence (e.g., recognition hypotheses) for multiple instances of the same utterance may be combined in a manner that produces a better estimate of the utterance.” (Franco, [0012].) Also, “a keyword such as ”repeating“ prefacing a subsequent utterance might indicate that the speaker is repeating an incorrectly recognized portion of the first utterance.” (Franco, [0015].) A speaker repeating all or part of the same utterance is an example of a “*predetermined event*” that an adaptive misrecognition engine would recognize. In both examples, the system reinterprets the misrecognized words from the first utterance in response to the predetermined event (the repetition of the utterance or use of a keyword by the user indicative of a recognition error) by using the first hypothesis in subsequent hypotheses. (Franco, [0020] (“the acoustic evidence from the previous and subsequent utterances can be combined in an efficient manner to produce a more accurate corrected recognition”).)

290. The reinterpretation in Franco is being performed by a “*conversational language processor*.” According to the ’176 patent specification, the conversational language processor builds “short-term shared knowledge” which “may include

information about previous utterances, requests, and other user interactions to inform generating an appropriate response to a current utterance.” (’176 patent, 4:17-23.) In addition, the conversational language processor “support[s] adaptive misrecognition to reinterpret a current utterance and/or one or more previous utterances.” (’176 patent, 4:45-47.) Because Franco discloses generating hypotheses for first and second utterances (at least one of which is part of “short-term shared knowledge”), Franco reinterprets the first utterance at a conversational language processor.

291. In my opinion, a POSITA would be motivated to incorporate the Franco misrecognition engine into the conversational language processor in Taschereau and Hansen (e.g., the multi-pass latent grammar in Taschereau or Command Processor in Hansen). As I discussed above, the structures in Taschereau and Hansen conduct recursive analyses similar to those in Franco as part of speech recognition. A POSITA would therefore be motivated to incorporate the Franco misrecognition engine into the conversational language processor in Taschereau and Hansen.

292. The combination of Taschereau, Hansen, and Franco thus discloses *“determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event”* as well as *“an adaptive misrecognition engine configured to determine that the conversational language*

*incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.”*

293. XX

**C. Dependent Claims 12 and 38**

12. The method of claim 1, further comprising determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

38. The system of claim 27, further comprising an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

294. Claim 12 and 38 depend from claim 1 and 27, respectively, and add the limitation “*further comprising determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.*”

295. The limitation added in claim 12 and 38 is the exact same limitation that is addressed in §IV.B.7 regarding claim limitation [14F] above. For the same

reasons discussed above in §VI.B.7 for claim limitation [14F], (i) the inventors admitted this limitation was well known in the prior art; and (ii) the combination of Taschereau and Hansen discloses “*determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.*”

**D. Dependent Claims 13, 15, 39 and 41**

13. The method of claim 12, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

15. The method of claim 14, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

39. The system of claim 38, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than

an expected time to process the request, or a subsequent input that repeats the natural language utterance.

41. The system of claim 40, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

296. Dependent claims 13, 15, 39 and 41 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 10, 25, 36 and 51 add the following limitation: “*wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.*”

297. As discussed above, I understand from the '176 patent that the inventors were aware that the user of an adaptive misrecognition engine was well known in the art and was not novel. ('176 patent, 4:55-64.) A POSITA would understand that the four types of “*predetermined events*” listed in this claim limitation (“*[i] a subsequent input proximate in time to the natural language utterance, [ii] a*

*subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, [iii] a subsequent input that overrides the request in a time shorter than an expected time to process the request, or [iv] a subsequent input that repeats the natural language utterance”*) are typical predetermined events that are regularly detected by adaptive misrecognition engines. Therefore, a POSITA would understand that the inventors conceded that this limitation was already well known in the prior art.

298. As discussed above related to claim limitation [14F], Franco discloses “*predetermined events*” that are detected by the “*adaptive misrecognition engine*” that cause the conversational language processor to reinterpret the words or phrases.

299. Franco discloses a predetermined event of “*a subsequent input that repeats the natural language utterance.*” Franco discloses a system that assumes “when speech recognition is unsuccessful on the first attempt, a speaker will **naturally repeat all or part of the same utterance in order to correct the recognition error.** Thus, acoustic evidence (e.g., recognition hypotheses) for multiple instances of the same utterance may be combined in a manner that produces a better estimate of the utterance.” (Franco, [0012].) Also, “a keyword such as ”repeating“ prefacing a subsequent utterance might indicate that the speaker is repeating an incorrectly recognized portion of the first utterance.” (Franco, [0015].)

A speaker repeating all or part of the same utterance is an example of a “*predetermined event*” that an adaptive misrecognition engine would recognize.

300. Franco further discloses a predetermined event of “*a subsequent input proximate in time to the natural language utterance.*” Specifically, the reference to “temporal intervals” in Franco demonstrates that the subsequent input must be proximate in time to the first utterance. As discussed above, Franco discloses that “[t]he method 200 is initialized at step 202 and proceeds to step 204, where the method 200 generates a first word lattice based on the first hypothesis. The first word lattice is a network or diagram that graphically illustrates the first hypothesis. Specifically, the first lattice may be thought of as comprising a plurality of lattice positions or word “slots“, where each slot corresponds to a temporal interval in the first utterance. Each slot is associated with one or more words that may potentially be have been uttered during the corresponding temporal interval. All of the slots in total comprise a phrase that may represent the first utterance (e.g., by selecting one word for each slot and then concatenating the selected words). Each potential word is associated with a score that indicates the probability that the word was in fact, uttered during the temporal interval corresponding to that slot. In one embodiment, a deletion may be the most probable word at one or more slots in the first word lattice. In fact, if the waveform representing the first utterance has a low signal-to-noise ratio, the first word lattice may include a plurality of slots whose most probable

word is a deletion, due to multiple insertions with low probabilities.” (Franco, [0022].) For example, a user may state: “[t]he number is 1234. I repeat, the number is 1234.” (Franco, [0018].) In Franco, “the method 100 would treat ”The number is 1234“ as the first utterance and ”I repeat, the number is 1234“ as the subsequent utterance, and combine respective first and subsequent hypotheses as described above to produce the combined hypothesis.” (Franco, [0018].) A POSITA would understand that Franco is describing a system where the “*conversational language processor reinterprets the words or phrases in response to [a] predetermined event,*” where the predetermined event is the system receiving subsequent inputs from the user that the system uses to reinterpret (or delete) prior utterances, meaning that “*the conversational language processor incorrectly interpreted the words or phrases.*”

301. The combination of Taschereau, Hansen and Franco discloses “*wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.*”

### **E. Dependent Claims 16 and 42**

302. Claims 16 and 42 include limitations [16P]-[16B] and [42P]-[42B]. These limitations are substantively identical to those in limitations [1C]-[1D]/[27C]-[27D]. For the same reasons discussed above in §V.B.4-5, the combination of Taschereau-Hansen-Franco combination discloses these limitations.

### **F. Dependent Claims 17, 18, 25, 26, 43, 44, 51, 52**

303. Claims 17 and 43 include the same limitations as in claims 9 and 35. Claims 18 and 44 include the same limitations as in claims 2 and 28. Claims 25 and 51 include the same limitations as in claims 10 and 36. Claims 26 and 52 include the same limitations as in claims 11 and 37. For the same reasons discussed above in §§V.D-F, the combination of Taschereau-Hansen-Franco combination discloses these limitations.

## **VII. GROUND 3: THE COMBINATION OF TASCHEREAU, HANSEN AND JONES RENDERS OBVIOUS CLAIMS 3-8, 19-24, 29-34 AND 45-50**

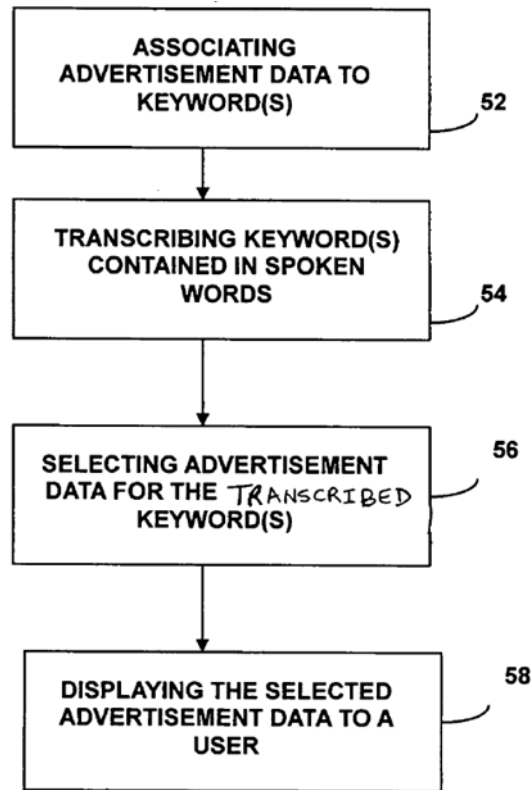
### **A. Overview of the Combination**

304. The combination of Taschereau, Hansen and U.S. Patent Application Publication No. US 2007/0174258 (“Jones”; EX-1008) discloses every limitation of claims 3-8, 19-24, 29-34 and 45-50. Taschereau, Hansen and Jones are analogous art to the ’176 patent because all three references are directed to speech recognition systems and software.

305. I have described both Taschereau and Hansen above, and I discuss Jones below.

# **1. Jones**

306. “The system and method disclosed [in Jones] provides targeted advertising to cellular platforms using voice recognition by identifying a keyword in a spoken phrase, looking up the keyword in an index having corresponding advertising and providing the advertising to cellular subscribers.” (Jones, [0066].) Specifically, “[t]argeted advertisement is provided to mobile device users” wherein “users may submit voice requests,” “relevant keyword(s) are determined from the requests and corresponding advertisements and/or products associated with the keyword(s) are provided to the users while the users are waiting for responses to the requests.” (Jones, Abstract.) Such features are depicted in Figure 2 shown below.



**Jones, Figure 2**

307. The system uses information associated with the user to determine what advertisement to provide, such as the user's phone number or the user's GPS location. "When a voice query from one of the user devices 12 is received by the system 10, for example, for a search without requesting assistance from a provider (or guide), the query is transcribed and passed to one or more search engines or resources, and advertisements associated with the query may be selected based on keyword histograms in the database 18 that 'best fit' the query and as appropriate for the user telephone number. For example, when a user requests information pertaining to 'restaurants in Indiana' and data records in the database 18 indicates

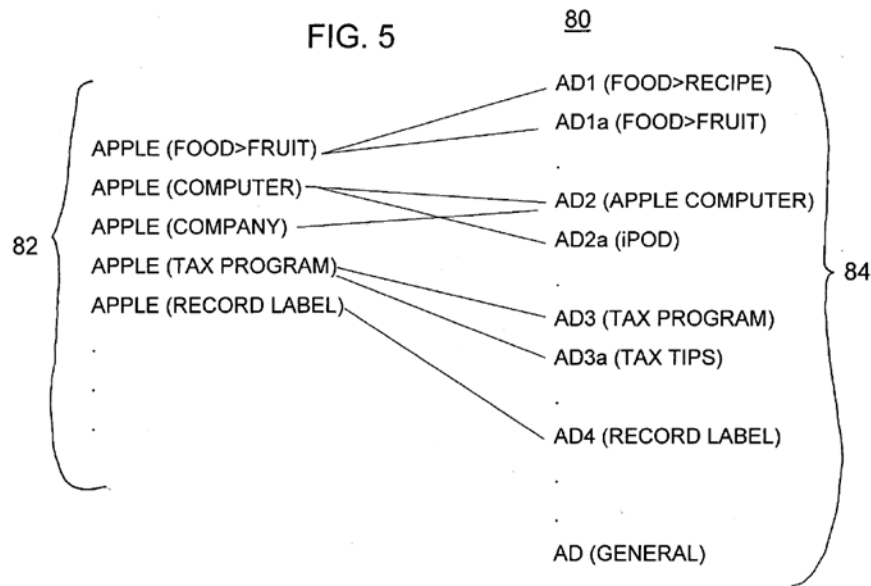
that a particular seafood restaurant advertisement has been presented to similar previous queries, the seafood restaurant advertisement may be presented to the user. The user telephone number or other geographic data such as GPS coordinates may also be used to select from restaurant advertisements in Indiana to provide one in or near the designated geographic location.” (Jones, [0031].)

308. To determine the appropriate advertisement to provide to users, the system has a database that maintains information about the user’s prior requests and what information was provided in response to those prior requests: “The database 18 may maintain information about ... requests submitted from the user devices 12, results generated by the providers in response to the requests, advertisements presented to the users while corresponding requests are being processed by the system 10 including advertisements previously presented for requests pertaining to similar subject matter.” (Jones, [0032].) The system also tracks whether users interact with certain advertisements after making requests with certain keywords: “For example, the database 18 may maintain a record of advertisements presented in response to queries containing particular keyword(s) and whether users interacted with the advertisements.” (Jones, [0032].)

309. The database also maintains information regarding user engagement with advertisements that is used by the system when selecting future advertisements: “For example, the server 16 may compute the number of times an advertisement has

been presented, the length of time a user (or requester) viewed an advertisement, whether the user ‘clicked through’ or accessed the advertisement, whether the user bought products or requested services from an advertiser’s website referred to in the advertisement (this information may be delivered back to the server 16 from the advertiser), etc.” (Jones, [0033].) The system may also select advertisements based on information about the user that is associated with the phone number and/or the user, i.e., a user profile. (Jones, [0063].)

310. The database also may maintain a histogram of usages for keywords or phrases with multiple meanings and/or forms. (Jones, [0053].) The system may prompt the user for clarification or “the system 10 may classify or categorize keywords and/or phrases based on information from external sources such as Internet directories (e.g., DMOZ), online dictionaries and/or encyclopedias (e.g., Webopedia, Wikipedia, etc.) or other similar sources.” (Jones, [0053].) Figure 5 shows the association of qualified keywords with advertisements:



**Jones, Figure 5**

311. However, the system may also use the context of the user’s search itself to determine the meaning of a phrase that may have ambiguity: “For example, advertisements associated with iPods may be selected for voice queries pertaining to apple products while advertisements associated with a particular type of the fruit apple may be selected for a query requesting information related to apple pie recipes.” (Jones, [0040].)

## **2. Motivation to Combine**

312. A POSITA would have been motivated to integrate the advertising analytic features disclosed in Jones into the combination of Taschereau and Hansen for a number of reasons.

313. First, Taschereau expressly provides motivation for the combination with Jones. Taschereau states that “[b]y selling this targeted advertising, it is possible for a service provider to provide directory assistance at a profit without charging users of the service for the calls. Given that the cost of the calls is a major constraint on the use of directory assistance services, by alleviating the cost, the demand for directory assistance will increase.” (Taschereau, 25:12-15.) A POSITA would have been motivated to incorporate the analytical features disclosed in Jones (e.g., tracking user interaction with the advertisement) in order to maximize the ability of the system to target advertisements to users and maximize revenue. Combining Jones with Taschereau provides a straightforward way for the Taschereau system to improve the advertisements it is selecting and providing to users, which will improve user experience and revenue.

314. Jones similarly discloses the goal of maximizing advertisement revenue and provides a variety of ways to monetize the advertising feature:

Income from advertisements may be based on conventional measures, such as person-minutes that an advertisement was heard, partially or completely with repetition of advertisements not counted or counted less, responses to ads based on selection input through the selection keys 104. For example, options for interacting with advertisements may be provided via the cell phone 100 that say “press ‘1’ if you want to hear more about this advertisement now, press ‘2’ if you want to hear more about this advertisement after receiving your search query results, press ‘3’ if you want us to send you a voice mail

message about a service/product included in this advertisement.”

(Jones [0062].) Jones discloses tracking user interaction so that income from advertisements can be based on that user interaction.

315. Second, Taschereau and Jones each also discuss the importance of being as specific as possible in targeting advertising to users of the telephone directory system. (Taschereau, 25:10-11 (“By using the information available about the user and the listing the user is looking for, very precise advertisements can be presented to the user.”); Jones, [0005] (“current advertising technologies are directed to mass advertisement and do not provide customized advertisements directed to interests particular to users or their needs”).)

316. A POSITA would understand that both Taschereau and Jones recognize that the more specific and accurate the advertisements are for a particular user, the better the user experience will be. In other words, a POSITA would understand that users prefer to engage with systems that are providing relevant advertisements over systems that are providing irrelevant advertisements. As the systems in Taschereau and Jones generate revenue from advertisement engagement, both systems are incentivized to maximize user experience.

317. Third, a POSITA would have had a reasonable expectation of success combining Taschereau, Hansen and Jones and the combination would have been predictable to a POSITA because both each of Taschereau, Hansen and Jones use

the same system architecture, namely, the use of ASR systems in conjunction with direct assistance systems.

**B. Dependent Claims 3, 19, 29 and 45**

**1. Preamble [3P], [19P], [29P] and [45P]**

- 3. The method of claim 2, further comprising:
- 19. The method of claim 18, further comprising:
- 29. The system of claim 28, wherein the conversational language processor is further configured to:
- 45. The system of claim 44, wherein the conversational language processor is further configured to:

318. The preambles for claims 3, 19, 29 and 45 simply refer back to the claims from which they depend and do not add any substantive limitations. As discussed above in §§V.C, VI.F, regarding claims 2, 18, 28 and 44, those claims are all disclosed in the prior art and would have been well known to a POSITA.

## 2. Tracking an Interaction Pattern [3A], [19A], [29A] and [45A]

[3A] tracking an interaction pattern with the advertisement presented via the output device; and

[19A] tracking an interaction pattern with the advertisement presented via the output device; and

[29A] track an interaction pattern with the advertisement presented via the output device; and

[45A] track an interaction pattern with the advertisement presented via the output device; and

319. The combination of Taschereau, Hansen and Jones discloses “*tracking an interaction pattern with the advertisement presented via the output device*” as well as conversational language processor is configured to “*track an interaction pattern with the advertisement presented via the output device.*” Specifically, Jones discloses these limitations.

320. Jones teaches that after an advertisement is presented to the user via the output device, the system tracks that interaction with the advertisement in a database:

For example, the server 16 may compute the number of times an advertisement has been presented, the length of time a user (or requester) viewed an advertisement, whether the user ‘clicked through’ or accessed the advertisement, whether the user bought products or requested services from an advertiser’s website referred to in the advertisement (this information may be delivered back to the server 16 from the advertiser), etc.

(Jones, [0033].) A POSITA would understand each of the following activities by Jones to be “*tracking an interaction pattern with the advertisement presented via the*

*output device*”: “[i] the length of time a user (or requester) viewed an advertisement, [ii] whether the user ‘clicked through’ or accessed the advertisement, [iii] whether the user bought products or requested services from an advertiser’s website referred to in the advertisement.” (Jones, [0033].)

321. Similarly, Jones teaches that the system updates the database to reflect if a user views, clicks or makes a purchase through an advertisement:

Further, if the user views an advertisement, or clicks on a link, or purchases a product or orders a service offered (sometimes referred to as “conversion”) related to an advertisement while awaiting results of a search, the database 18 may be updated to reflect additional ad revenue with a credit of points and/or compensation, if appropriate, to the provider (or searcher).

(Jones, [0028].)

322. Jones further teaches that geographical information is collected and stored by the system:

The guide may ask the user or an automated voice may ask the user for location information. The location information may be a zip code, a street number, a street intersection, a point of interest, latitude and longitude coordinates, an area code, or any other form of location information. ***This location information, once gathered by the system, may then be stored.*** Keywords may be linked to guides and/or advertisements that are tied to a specific geographic area that has a well-defined boundary.

(Jones, [0031].)

**3. Updating a Model Based on the Interaction Pattern [3B], [19B], [29B] and [45B]**

[3B] updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

[19B] updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

[29B] update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

[45B] update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

323. The combination of Taschereau, Hansen and Jones “*updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement*” as well as conversational language processor is configured to “*update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.*” Specifically, Jones discloses these limitations.

324. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other

characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

325. After the system described in Jones collects the “interaction pattern tracked for the advertisement,” it stores the information in the database as described above. The database includes the models which the system uses to select and advertisement, including an “*environmental model based on the interaction pattern tracked for the advertisement*”:

When a voice query from one of the user devices 12 is received by the system 10, for example, for a search without requesting assistance from a provider (or guide), the query is transcribed and passed to one or more search engines or resources, and **advertisements associated with the query may be selected based on keyword histograms in the database 18 that “best fit” the query and as appropriate for the user telephone number.**

(Jones, [0031].) Here, Jones is describing an “*environmental model based on the interaction pattern tracked for the advertisement*” because the system is selecting advertisements that best fit the query and the user’s telephone number, which indicates to the system where the user may be located and therefore interested in location specific information or advertisements.

326. Jones further discloses that that an advertisement having previous popularity with users and geographical location are used in the determining what advertisement to select and provide to a user:

An advertisement presented to a user may be based on a weight associated with the advertisement in accordance with a single or combined factors such as advertiser contract commitments, bidding price of advertisers, **popularity with users**, keyword mapping to advertisements, statistical usage (e.g. least recently presented), user demographics, provider choice of advertisement, **geographical location of the user** and/or provider, etc.

(Jones, [0029].)

327. According to the '176 patent, “an environmental model may be accessed to determine user location, user activity, track user actions, and/or other environmental information to invoke context, domain knowledge, preferences, and/or other cognitive qualities to enhance the interpretation of questions and/or commands.” ('176 patent, 4:64-5:3.) Therefore, a POSITA would understand the use of user location, user activity or tracking user actions would be the use of an “*environmental model*” pursuant to the claims.

328. By updating the database in Jones with “user location, user activity, [and] user actions,” the system described by Jones is “*updating ... the environmental model based on the interaction pattern tracked for the advertisement*” because user interactions with advertisements are tracked and used by the system to select and present advertisements in the future. For example, the system in Jones may associate keywords with advertisements that are tied to a specific geographic area and were previously interacted with by previous users:

For example, when a user requests information pertaining to “restaurants in Indiana” and data records in the database 18 indicates that a particular seafood restaurant advertisement has been presented to similar previous queries, the seafood restaurant advertisement may be presented to the user.

(Jones, [0031].)

### **C. Dependent Claims 4, 20, 30 and 46**

[4] The method of claim 3, wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

[20] The method of claim 19, wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

[30] The system of claim 29, wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

[46] The system of claim 45, wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

329. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or

idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

330. The combination of Taschereau, Hansen and Jones discloses “*wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances*” as well as “*wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.*” Specifically, Taschereau and Jones disclose these limitations.

331. As discussed above in §VII.A.1, Jones updates the database in the system to include certain types of information about user behavior or location, which are the “*environmental model.*” Also as discussed above, the system in Jones “*updates ... the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.*” Specifically, the database described in Jones, which includes the environmental model, builds statistical profiles for selecting subsequent advertisements for users. For example, Jones describes that advertisements may be selected by referencing a

keyword histogram, which is a type of statistical profile that is based on prior user interactions and location:

When a voice query from one of the user devices 12 is received by the system 10, for example, for a search without requesting assistance from a provider (or guide), the query is transcribed and passed to one or more search engines or resources, and advertisements associated with the query may be selected based on **keyword histograms in the database 18 that “best fit” the query and as appropriate for the user telephone number**. For example, when a user requests information pertaining to “restaurants in Indiana” and data records in the database 18 indicates that a particular seafood restaurant advertisement has been presented to similar previous queries, the seafood restaurant advertisement may be presented to the user.

(Jones, [0031].)

332. Also as discussed above, Jones discloses that an advertisement having previous popularity with users and geographical location are used in determining what advertisement to select and provide to a user:

An advertisement presented to a user may be based on a weight associated with the advertisement in accordance with a single or combined factors such as advertiser contract commitments, bidding price of advertisers, **popularity with users**, keyword mapping to advertisements, statistical usage (e.g. least recently presented), user demographics, provider choice of advertisement, **geographical location of the user** and/or provider, etc.

(Jones, [0029].)

333. A POSITA would understand that in order for Jones to know an advertisement is “popular with users” the system would necessarily have a “*build a statistical profile*” that indicates what advertisements are “popular” based on prior user interactions. For example, to determine if an advertisement is popular, the system in Jones would need to determine how often a user interacted with that particular advertisement compared to other advertisements, which would be represented as a statistical profile in the system (e.g., 40% of users interacted with advertisement A compared to 10% of users interacted with advertisement B).

334. Similarly, the system in Taschereau “updates ... the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.” Taschereau discloses that when a user makes a request, “[t]he user will also be providing information about themselves based on the area from which they are calling and the call display information.” (Taschereau, 25:8-9.) Taschereau then discloses that “[b]y using the information available about the user and the listing the user is looking for, very precise advertisements can be presented to the user.” (Taschereau, 25:10-11.) A POSITA would understand that in order for the Taschereau system to provide “very precise advertisements” to the user based on information about the user and the user’s location, the system would necessarily need to have an “*environmental model to build statistical profiles for selecting subsequent advertisements in response to*

*subsequent natural language utterances.”* The system would need to know statistically which advertisements are most likely to be relevant to a specific user’s location.

**D. Dependent Claims 5, 21, 31 and 47**

[5] The method of claim 4, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

[21] The method of claim 20, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

[31] The system of claim 30, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

[47] The system of claim 46, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

335. The combination of Taschereau, Hansen and Jones discloses “*wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action*

*associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.”* Specifically, Jones discloses this limitation.

336. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

337. As discussed above in §VII.A.1, the “*environmental model to build statistical profiles for selecting subsequent advertisements*” disclosed in Jones—a keyword histogram—is used to “*identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases*” to the system in Jones can select and present advertisements targeted advertisements to users. Those targeted advertisements are based on affinity between the recognized words or phrases uttered by the user and the advertisement that is presented via the output device:

When a voice query from one of the user devices 12 is received by the system 10, for example, for a search without requesting assistance from a provider (or guide), the query is transcribed and passed to one or more search engines or resources, and advertisements associated with the query may be selected based on keyword histograms in the database 18 that “best fit” the query and as appropriate for the user telephone number.

(Jones, [0031].)

338. Jones further discusses how the system is used to “*identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases*”:

The database 18 may also maintain information associated with a number of advertisements presented while voice queries are being processed by the system 10, **a number of times a particular advertisement is presented in association with one or more keywords**, advertisements hand-selected by providers, etc. For example, the server 16 may compute the number of times an advertisement has been presented, the length of time a user (or requester) viewed an advertisement, whether the user “clicked through” or accessed the advertisement, whether the user bought products or requested services from an advertiser’s website referred to in the advertisement (this information may be delivered back to the server 16 from the advertiser), etc.

(Jones, [0033].)

## **E. Dependent Claims 6, 22, 32 and 48**

### **1. Preamble [6P], [22P], [32P] and [48P]**

[6P] The method of claim 3, further comprising:

[22P] The method of claim 19, further comprising:

[32P] The system of claim 29, wherein the conversational language processor is further configured to:

[48P] The system of claim 45, wherein the conversational language processor is further configured to:

339. The preambles for claims 6, 22, 32 and 48 simply refer back to the claims from which they depend and do not add any substantive limitations. As discussed above in §V.B regarding claims 3, 19, 29 and 45, those claims are all disclosed in the prior art and would have been well known to a POSITA.

**2. Building Long-Term and Short-Term Knowledge [6A], [22A], [32A] and [48A]**

[6A] building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model;

[22A] building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model;

[32A] build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model; and

[48A] build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model; and

340. The combination of Taschereau, Hansen and Jones discloses “*building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model*” as well as a conversational language processor configured to “*build long-term shared knowledge and short-term shared knowledge*

*in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model.”* Specifically, Jones discloses these limitations.

341. As discussed above, Jones stores in the database information that it receives from users during interactions. For example, the system in Jones describes that “the database 18 may maintain a record of advertisements presented in response to queries containing particular keyword(s) and whether users interacted with the advertisements.” (Jones, [0032].)

342. The database also maintains information regarding user engagement with advertisements that is used by the system when selecting future advertisements: “For example, the server 16 may compute the number of times an advertisement has been presented, the length of time a user (or requester) viewed an advertisement, whether the user ‘clicked through’ or accessed the advertisement, whether the user bought products or requested services from an advertiser’s website referred to in the advertisement (this information may be delivered back to the server 16 from the advertiser), etc.” (Jones, [0033].) The system may also select advertisements based on information about the user that is associated with the phone number and/or the user, i.e., a user profile, which is long-term shared knowledge about the user that the system was aware of before the current conversation with the user began. (Jones, [0063].)

343. The system also builds “*long-term shared knowledge ... in response to updating ... the environmental model*” when the system builds long term knowledge about what types of advertisements users interact with in certain locations based on the user’s request:

For example, when a user requests information pertaining to “restaurants in Indiana” and data records in the database 18 indicates that a particular seafood restaurant advertisement has been presented to similar previous queries, the seafood restaurant advertisement may be presented to the user.

(Jones, [0031].) A POSITA would understand that the system described in Jones “*build[s] long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model.*”

344. Jones also discloses “*building short-term shared knowledge in response to updating ... the environmental model.*” For example, Jones discloses tracking a caller’s geographic information during a session. (Jones, [0026] (“The system 10 may keep track of the port from which a call originated as well as the telephone number of the caller, and assign user identifier to the user device 12 on this port for a particular session.”).) Such “geographical location of the user” is then used to select an advertisement to be presented. (Jones, [0029] (“For example, a voice query pertaining to a schedule of performances at The Kennedy Center for the Performing Arts in Washington, D.C. from a user who indicates being a resident of New York

or whose telephone number corresponds to a New York number, may trigger one or more advertisements related to Broadway shows in New York to be presented to the user, who might respond by purchasing tickets.”.) By leveraging the updated “environmental model” to select an advertisement, Jones discloses “*building short-term shared knowledge in response*” to the updated model.

345. Jones likewise describes that in some instances certain keywords or phrases have multiple meanings and the system must determine which meaning the user is referencing by building short term shared knowledge:

As mentioned above, the advertisement may be selected 72 by a provider or guide, or selected automatically by the system 10 (FIG. 1) based on a histogram that indicates classification or category of keywords or phrases identified by the system 10. The database 18 (FIG. 1) may maintain a histogram of usage for multiple meanings and/or forms of keyword(s) or phrase(s) which may be used to prompt a user (or consumer) or provider to qualify or clarify keyword(s) or phrase(s) having multiple meanings or forms when formulating a search query and an advertisement may be selected based on the qualified meaning of the keyword(s) or phrase(s). In addition, the system 10 may classify or categorize keywords and/or phrases based on information from external sources such as Internet directories (e.g., DMOZ), online dictionaries and/or encyclopedias (e.g., Webopedia, Wikipedia, etc.) or other similar sources. For example, a user (or provider) may be prompted to qualify keyword(s) contained in a query based on taxonomy utilized by DMOZ and an advertisement may be provided to the user based on meaning obtained from such qualification.

(Jones, [0053].)

346. Another example of Jones “*building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model*” relates to Jones using current geographic location information about the user that is obtained when the user makes the call. That short-term knowledge about the user’s location is used with the keyword analysis:

Any qualified keyword maintained in the database 18 (FIG. 1) may be mapped to advertisement information for presenting one or more targeted advertisement(s) to a user (or requestor) submitting a voice query containing the qualified keyword. For example, upon receipt of a query from a requester via a cell phone 12a (FIG. 1), for example, the server 16 may provide advertisement(s) to the requestor for viewing while a search is being processed by the system 10. As mentioned above, the advertisement information may be related to the query (or to the keywords of the query and/or the location of the telephone), may include visual, audio and/or interactive information as appropriate for the user’s device and for the source of the query and can include not only advertisements, but also information such as video, music, games, web links, etc. that will interact with and display data to the requestor while the search is being performed.

(Jones, [0059].)

### **3. Receiving a Subsequent Natural Language Utterance [6B] and [22B]**

[6B] receiving a subsequent natural language utterance at the input device; and

[22B] receiving a subsequent natural language utterance at the input device; and

347. The combination of Taschereau, Hansen and Jones discloses “*receiving a subsequent natural language utterance at the input device.*” Specifically, Jones discloses this limitation.

348. In the example described above in §VII.A.1, Jones discloses that in some instances the user utters a word that has multiple meanings. The system “may maintain a histogram of usage for multiple meanings and/or forms of keyword(s) or phrase(s)” and will “prompt a user (or consumer) or provider to qualify or clarify keyword(s) or phrase(s) having multiple meanings or forms when formulating a search query and an advertisement may be selected based on the qualified meaning of the keyword(s) or phrase(s).” (Jones, [0053].) When the system prompts the user to qualify or clarify a keyword or phrase, the system will then necessarily “*receiv[e] a subsequent natural language utterance at the input device.*” (See Jones, [0053].)

#### **4. Interpreting Using Long-Term and Short-Term Knowledge [6C], [22C], [32B] and [48B]**

[6C] interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.

[22C] interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.

[32B] interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.

|                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [48B] interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge. |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

349. The combination of Taschereau, Hansen and Jones discloses “*interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge,*” as well as a conversational language processor configured to “*interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.*” Specifically, Jones discloses this limitation.

350. In the example above in §VII.A.1, Jones discloses that the system “*interpret[s] the subsequent natural language utterance,*” which is the user’s qualification or clarification, using the conversational language processor described above in §VII.A.1, which uses long-term shared knowledge (e.g., information regarding overall user behavior in the current user’s geography or user interaction with advertisements) and short-term shared knowledge (e.g., the user’s current location) to interpret the “*subsequent natural language utterance.*” (Jones, [0053].) For example, if the user makes a “*subsequent natural language utterance*” that clarifies the initial natural language utterance, a POSITA would understand the system in Jones will necessarily “*interpret [the] subsequent natural language*

*utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.”*

**F. Dependent Claims 7, 23, 33 and 49**

[7] The method of claim 3, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

[23] The method of claim 19, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

[33] The system of claim 29, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

[49] The system of claim 45, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

351. The combination of Taschereau, Hansen and Jones discloses “*wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.*” Specifically, Jones discloses this limitation.

352. As discussed above in §VII.B.2, Jones discloses an “*tracking an interaction pattern with the advertisement,*” which is claimed in claim 3. Jones also discloses that “*the interaction pattern tracked for the advertisements*” includes “*an action performed in response to a subsequent request that identifies the*

*advertisement.*” Specifically, Jones discloses tracking when users click on advertisements or purchases products related to an advertisement:

Further, if the user views an advertisement, or clicks on a link, or purchases a product or orders a service offered (sometimes referred to as “conversion”) related to an advertisement while awaiting results of a search, the database 18 may be updated to reflect additional ad revenue with a credit of points and/or compensation, if appropriate, to the provider (or searcher).

(Jones, [0028].) In this example, a POSITA would understand that a subsequent request from a user identified an advertisement, which a user then may view, click on or make a purchase. When the user views, clicks on or makes a purchase that interaction pattern is tracked for the advertisement.

353. Jones provides additional disclosures regarding instances when “click-through” engagements or purchases come after the user provides a subsequent request that identifies the advertisement. For example, Jones teaches that after the system provides an advertisement, the system will provide the user an option of receiving a coupon for related to the advertisement. If the user subsequently requests the coupon, the system will provide the user the coupon:

Subsequent to delivering 74 the advertisement, the process 70 moves to providing 76 an option to receive a coupon for the goods/services indicated in the advertisement. Using the same example in the previous discussion, if a user interacts with the advertisement pertaining to the specific discounted Dell products, an option to receive a coupon for the products may be provided to the user (see also FIG. 3B). In the case of electronic merchandise (eg.

MP3 downloads) a product may be delivered in addition to an advertisement. For example the advertisement may contain video or audio ‘clips’ and the consumer may elect to purchase the full-length product, which is then supplied to him.

When a user selects the option provided 76 for receiving a coupon, process 70 moves to sending 78 a text message which may be sent to the user using, for example, Short Message Service (SMS) or other similar service. For example, a coupon for a 15% discount at a particular restaurant may be sent to a user’s cell phone in response to a voice query pertaining to the restaurant or other similar restaurants. When the mobile user device has image display capability, an image of a coupon may be sent.

(Jones, [0055]-[0056].) A POSITA would understand the user requesting a coupon related to the advertisement to be “*a subsequent request that identifies the advertisement.*” In response, the user may interact with the coupon (“an action”) and the system will track that “*interaction pattern.*”

354. Jones provides additional information regarding how user’s interaction pattern is tracked for advertisements:

The database 18 may also maintain information associated with a number of advertisements presented while voice queries are being processed by the system 10, **a number of times a particular advertisement is presented in association with one or more keywords**, advertisements hand-selected by providers, etc. For example, the server 16 may compute the number of times an advertisement has been presented, the length of time a user (or requester) viewed an advertisement, whether the user “clicked through” or accessed the advertisement, whether the user bought products or requested services from an advertiser’s

website referred to in the advertisement (this information may be delivered back to the server 16 from the advertiser), etc.

(Jones, [0033].)

#### **G. Dependent Claims 8, 24, 34 and 50**

[8] The method of claim 7, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

[24] The method of claim 23, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

[34] The system of claim 33, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

[50] The system of claim 49, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

355. The combination of Taschereau, Hansen and Jones discloses “*wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.*” Specifically, Jones discloses this limitation.

356. As discussed above in §VII.A.1, Jones teaches that the system “*execut[es] a task,*” by providing a coupon “*based on the subsequent request that identifies the advertisement,*” the subsequent request being the user’s request for a coupon that identifies the advertisement. (Jones, [0055]-[0056].)

## **VIII. GROUND 4: THE COMBINATION OF RAMER AND CROSS RENDERS OBVIOUS CLAIMS 1- 52**

### **A. Overview of the Combination**

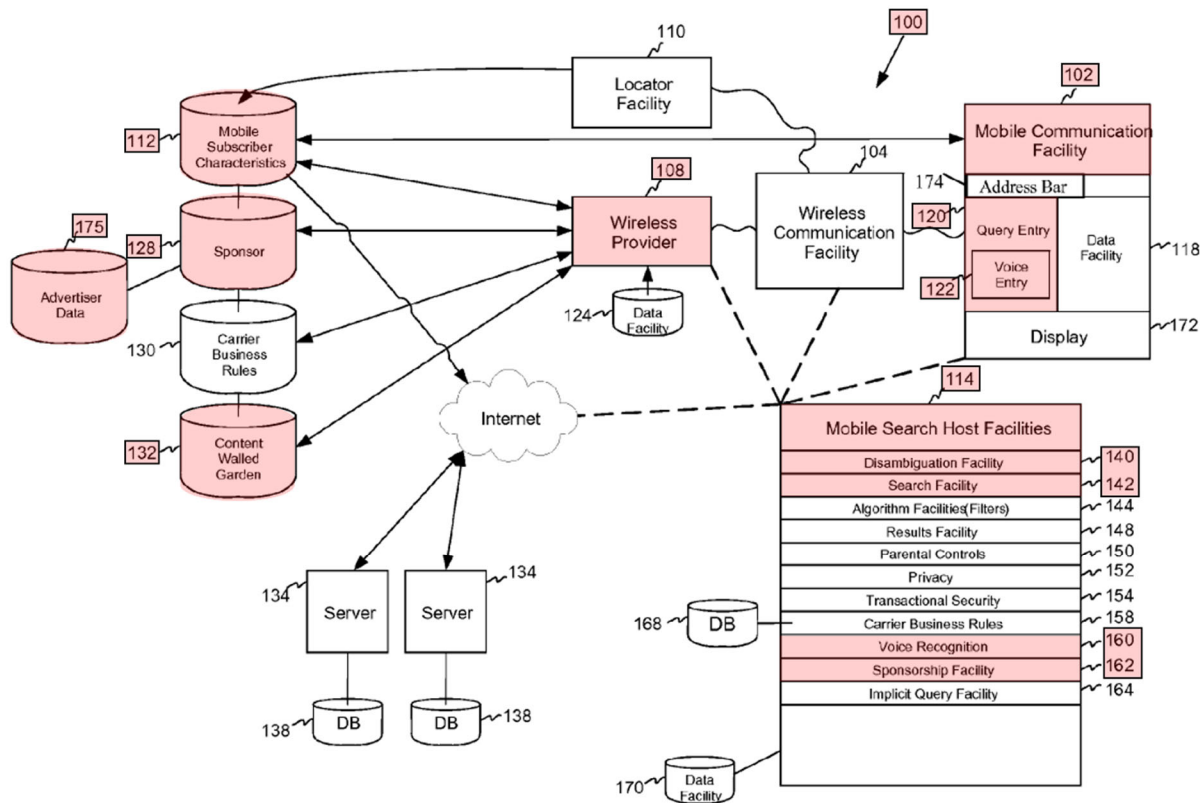
357. The combination of U.S. Patent No. 7,577,665 to Ramer, et al. (“Ramer”; EX-1009) and U.S. Patent 8,332,218 to Cross et al. (“Cross”; EX-1010) discloses every limitation of claims 1-52. Ramer and Cross are analogous art to the ’176 patent because both references are directed to speech recognition systems and software.

358. I describe both Ramer and Cross further below.

#### **1. Ramer**

359. Ramer discloses the use of speech recognition on a device to facilitate internet searches as well as selecting and presenting an advertisement to a user conducting a search based on past search history and other information about the user: “Information relating to the user characteristics associated with a mobile communication facility and other capabilities are employed to improve the presentation and relevance of mobile content to appropriate or desirable mobile communication facilities.” (Ramer, Abstract.)

360. Ramer “relates to the field of mobile communications and more particularly to improved mobile advertisement syndication for mobile communication facilities.” (Ramer, 1:47-49.) Figure 1 in Ramer is an exemplary “wireless search platform.” (Ramer, 5:6-7.)



**Ramer, Figure 1**

361. “[T]he mobile communication facility 102 may be any device capable of wireless communication, including, but not limited to a mobile phone, cell phone, satellite phone, walkie-talkie, handheld device, personal digital assistant (PDA), mobile network appliance, 6r an email, instant messaging, or chat device.” (Ramer, 5:22-27.) Data may be entered on the “mobile communication facility” through “query entry 120” (e.g., a keyboard, photographic entry, stylus, etc.) or “voice entry 122.” (Ramer, 5:49-51.) “The voice entry 122 function may permit voice dialing, voice memo, voice recognition, speech recognition, or other functions related to audible input.” (Ramer, 6:24-26.)

362. “The wireless search platform 100 illustrated in FIG. 1 may contain a data facility containing mobile subscriber characteristics 112 pertaining to individual users of a mobile communication facility 102. This data may include, but is not limited to, data collected by the wireless provider 108 when an individual opens a wireless account .... The mobile subscriber characteristics facility 112 may continually, or periodically, update data for individual users, for example, bill amount(s), average bill total, payment history, on-time payment history, on-line usage amount, duration of on-line interactions, number of on-line interactions, ... shopping habits (e.g., views of or purchases of goods and services)[,] click stream information, device type and device version, device characteristics, usage patterns (including those based on location, time of day, or other variables), device and/or subscriber unique identifiers, content viewing history, content presented for [sic] viewed by/not viewed by user, ... transaction history, and any other user or user defined characteristics.” (Ramer, 8:21-48.)

363. Ramer likewise discloses an advertising database for presenting sponsored content to users. (Ramer, 8:66-9:3 (“FIG. 1 illustrates a sponsorship facility 175 associated with a sponsor database 128 according to the principles of the present invention. The sponsorship facility 175 may be provided by a corporation, an individual, or some other entity sponsoring results as described herein.”).) “[T]he sponsor database 128 may store sponsor information in an accessible form in the

sponsor database 128 to be used in the searching for information, presentation of information, accessing of information, or other activity associated with the mobile communication facility 102. In embodiments, the sponsor information may be used to present syndicated sponsor links, advertising, content, or other information on the mobile communication facility 102.” (Ramer, 18:34-42.)

364. “The wireless search platform 100 may include mobile search host facilities 114. The mobile search host facilities 114 may include one or more facilities for disambiguation 140, searching 142, algorithms/filters 144, results 148, parental controls 150, privacy 152, transactional security 154, carrier business rules 158, voice recognition 160, sponsorship 162, and/or implicit query 164, either alone or in combination.” (Ramer, 4:47-54.)

365. “A voice recognition facility 160 may be a software component enabling a machine or device (e.g., a cellular phone) to understand human spoken language and to carry out spoken commands. Typically, a human voice is received by the device and converted to analog audio. The analog audio may in turn be converted into a digital format using, for example, an analog-to-digital converter, which digital data may be interpreted using voice recognition techniques. Generally this is done 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.” (Ramer, 21:22-33.)

366. “An implicit query facility 164 may automatically generate context-sensitive queries based on a user’s current activities, characteristics, and/or the user’s device characteristics. For example, the implicit query facility 164 may retrieve Internet links, music files, e-mails, and other materials that relate to an active user’s query, but which the user did not specifically query.” (Ramer, 21:46-52.)

367. “The search facility 142 may initiate a search, such as by causing a query (optionally a disambiguated query) to be executed on a search facility, such as a search engine.” (Ramer, 9:28-30.)

368. “FIG. 2 is a schematic diagram 200 showing a plurality of processes for handling a user query and producing a delivered result to the user.” (Ramer, 15:33-35.)

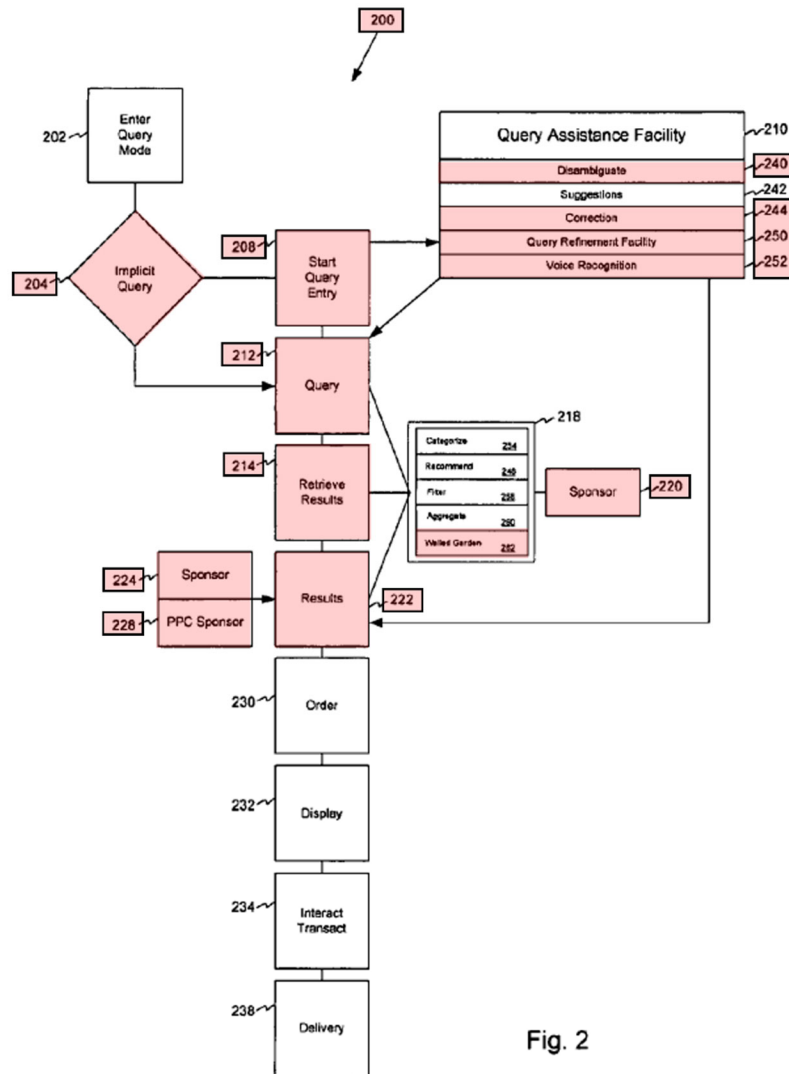


Fig. 2

### Ramer, Figure 2

369. “The start of the query entry 208 made explicitly by the user may be paired with additional information derived from a related implicit query 204. Depending on the clarity of the query entry 208, the query may need correction 244, disambiguation 240, or redirection 250. ... If the start of the query entry 208 is made by voice, rather than text, it may be aligned with a voice recognition 252 program.” (Ramer, 15:40-48.)

370. “Once the initial explicit and/or implicit query is made, the query 212 is processed and the initial results retrieved 214. ... Walled garden content 262 and sponsored content 220 may also attach to the query 212 and present tailored results 222 to the user. The results 222 may also, in turn, trigger the posting of additional sponsor 224 messages and advertisements. Information from pay-per-click (PPC) sponsors 228 may link to the results, making it possible for the user to quickly learn about sponsors’ services, phone numbers, addresses, hours of operation, sales, and so forth.” (Ramer, 15:53-64.) “Walled garden content 132 may be content from which the wireless provider 108 derives additional revenues based, for example, on user clickthroughs or content downloads.” (Ramer, 15:24-26.) For example, specific advertisements may be presented to a user based on past search history. (Ramer, 46:18-26 (“the bookstore in a building 1102B may want to advertise to the user 1104 because he is in the area and has some history of searching amazon.com and the like”).)

371. While Ramer discloses receiving a natural language utterance, recognizing one or more words or phrases in the natural language utterance, interpreting the natural language utterance, including establishing a context for the natural language utterance, and selecting and presenting an advertisement based on such context, it does not disclose details for how the speech recognition system performs such steps. A POSITA would have looked to known ways of implementing

such a speech recognition system. One well-known way of conducting such analysis is to map a stream of phonemes contained in the natural language utterance as taught in, for example, Cross.

## **2. Cross**

372. Cross is titled “Context-Based Grammars for Automated Speech Recognition.” The Abstract explains the patent teaches “[m]ethods, apparatus, and computer program products for providing a context-based grammar for automatic speech recognition, including creating by a multimodal application a context, the context comprising words associated with user activity in the multimodal application, and supplementing by the multimodal application a grammar for automatic speech recognition in dependence upon the context.” (Cross, Abstract.)

373. “A system that provides a context-based grammar for automatic speech recognition according to embodiments of the present invention augments existing speaker independent voice recognition engines by dynamically providing grammars, where they are otherwise unavailable, based on the users previous context of activity on the system.” (Cross, 17:39-45.)

374. “A multimodal device is an automated device, that is, automated computing machinery or a computer program running on an automated device, that is capable of accepting speech input from a user, digitizing speech, and providing digitized speech to an ASR engine for recognition.” (Cross, 3:24-28.) Exemplary

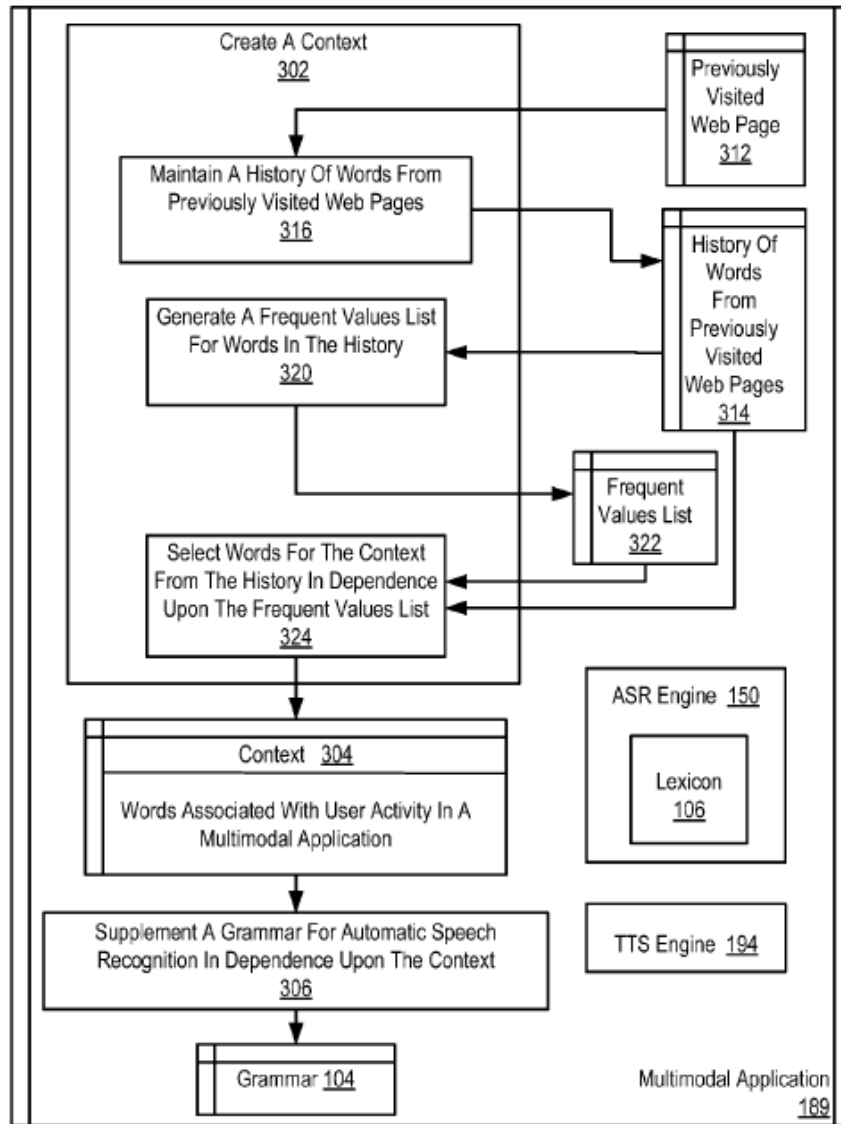
multi-modal devices include personal computers, PDAs, mobile telephones and laptop computers. (Cross, 3:34-47.) “An ASR engine (148) may be installed locally in the multimodal device itself, or an ASR engine (150) may be installed remotely with respect to the multimodal device, across a data communications network (100) in a voice server (151).” (Cross, 4:32-35.)

375. The ASR engine “is a module of computer program instructions ... that includes an ASR lexicon (106) of words capable of recognition by the ASR engine.” (Cross, 6:28-30.) “The lexicon (106) is an association of words in text form with phonemes representing pronunciations of each word.” (Cross, 6:30-32.) “[T]he ASR engine receives speech for recognition in the form of at least one digitized word, uses frequency components of the digitized word to derive a Speech Feature Vector (‘SFV’), uses the SFV to infer phonemes for the word from a language-specific acoustic model.” (Cross, 6:33-38.) The acoustic model “associates SFVs with phonemes representing, to the extent that it is practically feasible to do so, all pronunciations of all the words in a particular language. The ASR engine then uses the phonemes to find the word in the lexicon.” (Cross, 6:38-43.)

376. The system then conducts a contextual analysis to confirm that the word generated by the ASR engine is the right one. (*See, e.g.*, Cross, 12:56-64 (“multimodal device application (195) in this example is also configured to provide a context-based grammar for automatic speech recognition ... by creating by a

multimodal application a context, the context comprising words associated with user activity in the multimodal application, and supplementing by the multimodal application a grammar for automatic speech recognition in dependence upon the context”).)

377. “A grammar communicates to an ASR engine the words and sequences of words that may be recognized.” (Cross, 6:44-46.) Figure 8 illustrates an “exemplary method of providing a context-based grammar for automatic speech recognition.” (Cross, 14:48-51.)



**Cross, Figure 8**

378. “[C]reating (302) a context (304) is carried out by maintaining (316) a history (314) of words from previously visited web pages (312), generating (320) a frequent values list (322) for words in the history, and selecting (324) words for the context from the history in dependence upon the frequent values list.” (Cross, 14:59-64.) For example, a user who had conducted specific searches regarding an Acura

vehicle (e.g., “Compare the Acura RL and the BMW Z4” or “Find an Acura Dealer in Palm Beach, Fla.”) would generate a web history that would permit a “generic search page [to] recognize a phrase such as ‘Look for Acura roof racks.’” (Cross, 17:15-32.)

### 3. Motivation to Combine

379. A POSITA would have been motivated to combine the teachings of Cross regarding the specific details of context-based speech recognition, including mapping phonemes to syllables phonemically represented in an acoustic grammar, with the speech recognition driven content delivery system disclosed in Ramer.

380. A POSITA would have been motivated to make this combination for numerous reasons. First, Ramer suggests the combination. Ramer states:

A voice recognition facility 160 may be a software component enabling a machine or device (e.g., a cellular phone) to understand human spoken language and to carry out spoken commands. Typically, a human voice is received by the device and converted to analog audio. **The analog audio may in turn be converted into a digital format using, for example, an analog-to-digital converter, which digital data may be interpreted using voice recognition techniques.** Generally this is done 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. The speech patterns of a unique user may be stored on a hard drive (locally or remotely) or other memory device, and may be loaded into memory, in whole or in part, when the program is run. A comparator may use, for example, correlation or other discrete Fourier

transform or statistical techniques to compare the stored patterns against the output of the analog-digital converter.

The capacity of a voice recognition facility 160, such as vocabulary, speed, and length of digital samples that can be analyzed, may be constrained by hardware capabilities of the voice recognition facility 160, such as memory capacity, sampling rates, and processing speed, as well as the complexity of algorithms used for comparisons.

(Ramer, 21:22-45.)

381. Ramer instructs that the user utterances (audio) may be interpreted “using voice recognition techniques” which a POSITA would understand to mean prior art voice recognition techniques like those disclosed in Cross.

382. Second, Cross acknowledges that its speech recognition system is well known. (*See* Cross, 9:44-49 (discussing commercially available speech recognition systems which can be modified to add the contextual analysis set forth in Cross).)

383. Finally, the combination is nothing more than the application of a known technique (the well-known process for implementing speech recognition) to a similar known system (Ramer’s speech recognition system that facilitates its search and targeted advertisement capability) that operate in the same way (analyzing utterances through investigation of specific words and dynamic development of grammars). The system of Ramer is ready for improvement and the simple substitution of one known element (Cross’s ASR using phoneme mapping) for another (Ramer’s agnostic ASR). The results of the combination would have

been predictable as I discuss below. And, a POSITA would have been motivated to make the combinations for the reasons discussed above.

384. A POSITA would have had a reasonable expectation of success in the combination and the results of the combination would have been predictable. Combining Ramer with Cross would merely integrate one component (the context based ASR techniques) into Ramer’s system of receiving user utterances and returning targeted search results and advertisements using information about the user. Ramer expressly states that its voice recognition techniques can be used in the system. (*See* Ramer, 21:22-45.) Cross provides an example of a well-known processing technique that can be used in the Ramer system. (*See* 9:44-49.) Integrating Cross’s context-based speech recognition and processing into Ramer’s system to improve the processing of utterances would have been well within the skill set of a POSITA before the earliest priority date of the ’176 patent.

385. Ramer and Cross are both in the same field of the ’176 patent—using speech recognition to facilitate access to information (e.g., data or files) or programs. (*See* ’176 patent, 1:66-2:2 (“a system and method for selecting and presenting advertisements based on natural language processing of voice-based inputs is provided”); Ramer, 1:47-49 (“This disclosure relates to the field of mobile communications and more particularly to improved mobile advertisement syndication for mobile communication facilities.”), 3:57-59 (“The methods and

systems disclosed herein relate to the domain of mobile communication facilities and to the domain of searches for content.”).)

## **B. Independent Claims 1 and 27**

386. Independent claims 1 and 27 include substantially overlapping processing limitations ([1A]-[1G]/[27A]-[27G]), as shown in the table below. The only substantive difference between the claims is that claim 1 claims a method for selecting and presenting advertisements and claim 27 claims a system that practices that same method for selecting and presenting advertisements. Because of the overlap the limitations in these claims, I address these limitations together. In this table, I also provide a citation to the section discussing each limitation:

| <b>Claim 1</b>                                                                                                                                                               | <b>Claim 27</b>                                                                                                                                                                                                | <b>Section</b> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| [1P] A method for selecting and presenting advertisements in response to processing natural language utterances, comprising:                                                 | [27P] A system for selecting and presenting advertisements in response to processing natural language utterances, comprising                                                                                   | §VIII.B.1      |
| [1A] receiving a natural language utterance containing at least one request at an input device;                                                                              | [27A] an input device that receives a natural language utterance containing at least one request at an input device;                                                                                           | §VIII.B.2      |
| [1B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or | [27B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words | §VIII.B.3      |

| <b>Claim 1</b>                                                                                                                                                                                                                                              | <b>Claim 27</b>                                                                                                                                                                                                                                                                                                         | <b>Section</b> |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| phrases in the natural language utterance includes:                                                                                                                                                                                                         | or phrases in the natural language utterance, the speech recognition engine is configured to:                                                                                                                                                                                                                           |                |
| [1C] mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and                                                                                        | [27C] map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and                                                                                                                                                       | §VIII.B.4      |
| [1D] generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;                   | [27D] generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;                                                                                | §VIII.B.5      |
| [1E] interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | [27E] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | §VIII.B.6      |
| [1F] selecting an advertisement in the context established for the natural language utterance; and                                                                                                                                                          | [27F] select an advertisement in the context established for the natural language utterance; and                                                                                                                                                                                                                        | §VIII.B.7      |

| Claim 1                                                                                                           | Claim 27                                                       | Section   |
|-------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|-----------|
| [1G] presenting the selected advertisement via an output device coupled to the conversational language processor. | [27G] present the selected advertisement via an output device. | §VIII.B.8 |

# **1. Preamble [1P] / Preamble [27P]**

|                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>[1P] A method <b>for selecting and presenting advertisements in response to processing natural language utterances, comprising:</b></p> <p>[27P] A system <b>for selecting and presenting advertisements in certain, comprising</b></p> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

387. To the extent the preamble is limiting, the combination of Ramer and Cross discloses “[a] *method for selecting and presenting advertisements in response to processing natural language utterances*” as well as a “*system for selecting and presenting advertisements in response to processing natural language utterances,*” as discussed below in §§VIII.B.1-8. Ramer discloses a system that processes natural language utterances to assist users with searching and teaches that “recommended advertisements may be paired with webpages presented prior to, during, or following, the display of the advertisement results.” (Ramer, 2:34-36.) As discussed below, the combination of Ramer and Cross discloses a “*system*” that practices a “*method*” “*for selecting and presenting advertisements in response to processing natural language utterances,*” so I analyze these claims together.

388. Ramer discloses a method and system for selecting and presenting advertisements in “*response to processing natural language utterances*” as it discloses that a “search query may be entered on a mobile communication facility 102 using voice recognition and retrieve results based on the search query.” (Ramer, 83:59-61.)

389. Ramer further discloses that the user’s natural language utterance is used for “*selecting and presenting advertisements*”: “In embodiments, an auction for search marketing may be performed related to a presentation of sponsored content on a mobile communication facility 102, where the auction is the result of a query processed through a voice recognition facility 160 residing on the mobile communication facility 102.” (Ramer, 97:52-56.)

390. Ramer also discloses that the selecting and presentation of advertisements is based on prior user tracking: “In embodiments, a search query may be processed through a voice recognition facility 160 and a recommendation made in association with information about a mobile communication facility 102. For example, the voice recognition facility 160 may be employed to search for a particular restaurant. Based upon previous calls to other restaurants, a recommendation may be made for alternative restaurants of the same food genre in the users area, or a suggestion may be made that the user is trying to find a restaurant with a similar sounding name that the user has called before.” (Ramer, 72:28-37.)

391. Ramer further discloses that the system may provide the advertisement as a sponsored link or as part of the search results provided to the user: “In embodiments, a syndicated advertising result may be produced on a mobile communication facility 102, where the syndicated advertising result is produced in association with a query entered through a voice recognition facility 160. The syndicated advertising result may be displayed on a mobile communication facility 102 as a sponsored link. Syndicated advertising may be placed in, and prioritized within, the result set on the basis of an auction among sponsors.” (Ramer, 110:9-16.)

392. Ramer also discloses that the user engagement with advertisements is tracked and that the advertisements may be presented in various formats: “In embodiments, a voice recognition facility 160 may be associated with a sponsor facility and a mobile communication facility 102, and the voice recognition facility 160 may be used in the process of obtaining search results. Sponsor results may be paid inclusion results, auction results, or pay-per-click results (in connection with a WAP site or a phone number). A sponsor of the sponsored result may receive compensation as a result of activity associated with a mobile communication facility 102 phone number. The presentation of the sponsored results may be formatted as a link or presented as text, as a picture, as a video, or as an interactive application.” (Ramer, 120:46-57.)

393. The combination of Ramer and Cross discloses an “[a] *method for selecting and presenting advertisements in response to processing natural language utterances*” and “[a] *system for selecting and presenting advertisements in response to processing natural language utterances.*”

## 2. Receiving Natural Language Request [1A] / [27A]

[1A] receiving a natural language utterance containing at least one request at an input device;

[27A] an input device that receives a natural language utterance containing at least one request at an input device;

394. The combination of Ramer and Cross discloses “*receiving a natural language utterance containing at least one request at an input device*” and “*an input device that receives a natural language utterance containing at least one request at an input device.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

395. The ’176 patent describes a “request[]” as “performing an action, retrieving information, or various combinations thereof.” (’176 patent, 4:1-3.) The ’176 patent also provides examples of types of “requests”:

For example, content/action identification module 235 may identify content of the voice-based input (e.g., words in the input), requested information (e.g., search results, a web page, music, video, graphics, or other information), requested actions (e.g., calculating a navigation route, placing a telephone call, playing a song, etc.), a category or topic related to the input (e.g., music, business, stocks,

sports, navigation, movies, etc.), or other criteria to use in selecting an advertisement.

('176 patent, 5:56-6:1)

396. Ramer discloses a system that receives natural language utterances containing at least one request, like those described in the '176 patent. Ramer discloses that “[a] content request may be made by an explicit query request entered by a mobile communication facility user as a voice command in the voice recognition query facility of a mobile communication facility.” (Ramer, 2:23-26.) Exemplary “mobile communication facilit[ies]” include mobile phones. (Ramer, 5:22-27.) The mobile phone is an example of an “*input device*.”

397. Cross likewise discloses this limitation. Examples of “*natural language utterances utterance containing at least one request*” discussed in Cross include:

“Show me a quote for an RL”

“Compare the Acura RL and the BMW Z4”

“Find an Acura Dealer in Palm Beach, Florida”

(Cross, 16:50-52.) Like Ramer, such utterances can be received by mobile telephones, (Cross, 3:34-47), a type of “*input device*.”

398. The combination of Ramer and Cross discloses an “*receiving a natural language utterance containing at least one request at an input device*.”

### 3. Recognizing One or More Words in the Utterance [1B] / [27B]

[1B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or phrases in the natural language utterance includes:

[27B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to:

399. The combination of Ramer and Cross discloses “*recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device, wherein recognizing the words or phrases in the natural language utterance includes*” and “*a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

400. While Ramer discloses the use of voice recognition to facilitate internet search on a mobile device, it does not provide technical details as to how the voice recognition works. Cross provides those details.

401. Cross discloses “*recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input*

*device.*” Cross is titled “Context-based Grammars for Automated Speech Recognition.” The Cross Abstract explains that “[m]ethods, apparatus, and computer program products for providing a context-based grammar for automatic speech recognition, including creating by a multimodal application a context, the context comprising words associated with user activity in the multimodal application, and supplementing by the multimodal application a grammar for automatic speech recognition in dependence upon the context.” (Cross, Abstract.)

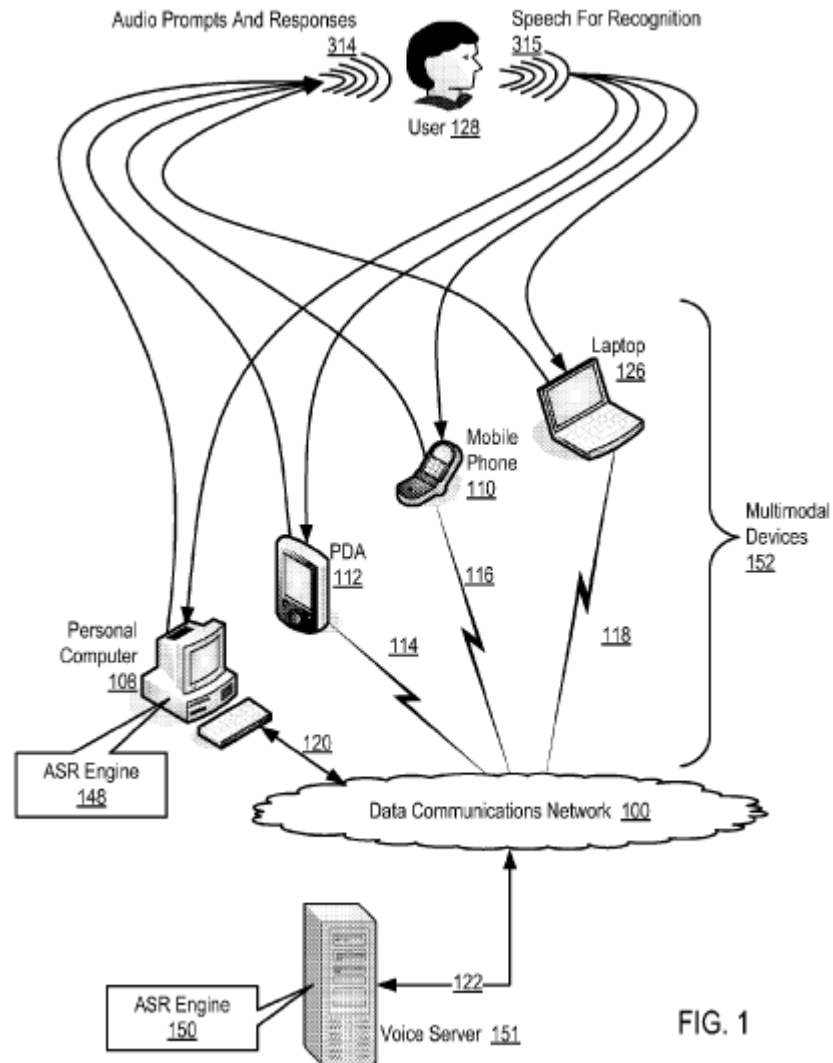
402. Cross further discloses:

The disclosed invention seeks to overcome the limitations of the current state of the art in speaker independent voice recognition by dynamically building voice recognition grammars based on the context of a user’s activities in a multimodal application. Methods, apparatus, and computer program products are described for providing a context-based grammar for automatic speech recognition, including creating by a multimodal application a context, the context comprising words associated with user activity in the multimodal application, and supplementing by the multimodal application a grammar for automatic speech recognition in dependence upon the context.

(Cross, 2:6-16.)

403. Specifically, Cross discloses that a “multimodal server application (188)” “receive[s] speech for recognition from a user and pass[es] the speech along to an ASR engine (150).” (Cross, 6:26-28.) “The ASR engine (150) contains a grammar (104). A grammar communicates to an ASR engine the words and sequences of words that may be recognized.” (Cross, 6:44-46.) Cross thus discloses

a “*speech recognition engine*” under the Amazon-VBA-Litigation construction. As shown in Figure 1 below, the ASR engine is “*coupled to the input device*” (e.g., mobile phone 110).



**Cross, Figure 1**

404. The combination of Ramer and Cross discloses an “*recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input devices.*”

#### 4. Mapping Phonemes [1C] / [27C]

[1C] mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

[27C] map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

405. The combination of Ramer and Cross discloses “*mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar*” and “*map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

406. As discussed above, I understand that in the prior District of Delaware proceeding the term “acoustic grammar” was construed to mean “grammar of phonotactic rules of the English language that maps phonemes to syllables.” (Amazon Litigation Mem. Op. Regarding Claim Construction, 2.)

407. I further understand that in the prior IPR proceeding the Board interpreted acoustic grammar as follows:

Thus, Kennewick ’409 explains that an “acoustic grammar” is a collection of the phonemes, or distinct units of sound of a spoken language, linked together to form syllables, which are linked together to form the words of the language. On this record and for the purposes of this

Decision, we interpret “acoustic grammar” as used in the ’176 patent in the same manner.

(Amazon-VBA-IPR Decision, 12.) It is my opinion that this definition of “acoustic grammar” is consistent with the definition used in the District of Delaware proceeding. I have applied these definitions to my opinions here.

408. While Ramer discloses the use of voice recognition to facilitate internet search on a mobile device, it does not provide technical details as to how the voice recognition works, such as “*mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar.*” Cross provides those details.

409. Cross discloses that “the ASR engine receives speech for recognition in the form of at least one digitized word, uses frequency components of the digitized word to derive a Speech Feature Vector (‘SFV’), uses the SFV to infer phonemes for the word from a language-specific acoustic model.” (Cross, 6:33-38.)

410. Cross provides additional information regarding the ASR engine: “ASR engine (150) is a module of computer program instructions, also stored in RAM in this example, that includes an ASR lexicon (106) of words capable of recognition by the ASR engine. The lexicon (106) is an association of words in text form with phonemes representing pronunciations of each word.” (Cross, 6:28-33.)

411. Cross further discloses that “[a] language-specific acoustic model is a data structure, a table or database, for example, that associates SFVs with phonemes

representing, to the extent that it is practically feasible to do so, all pronunciations of all the words in a particular language.” (Cross, 6:38-43.)

412. A POSITA would understand the reference to “pronunciations” in Cross to be synonymous with syllables. Because the model includes “all pronunciations of all the words in a particular language,” the stream of phonemes is mapped to one or more syllables that are phonemically represented in an acoustic grammar. As I discussed above (*supra* §V.B.4), and as discussed in Verma and Jones ’330, phonetic dictionaries that mapped phonemes to syllables were well known in the art. Moreover, Cross discloses that phonemes are generated “from a language-specific acoustic model,” (Cross, 12:1-6), such as English.

413. Cross thus discloses an “*acoustic grammar*” because it includes a grammar of phototactic rules of the English language that maps phonemes to syllables under the Amazon-VBA-Litigation construction and a collection of the phonemes linked together to form syllables, which are linked together to form the words of the language under the Amazon-VBA-IPR interpretation.

414. The combination of Ramer and Cross thus discloses an “*mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar.*”

## 5. Generating Preliminary Interpretation of Utterance [1D] / [27D]

[1D] generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;

[27D] generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases;

415. The combination of Ramer and Cross discloses “*generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases*” and a speech recognition engine configured to “*generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

416. While Ramer discloses the use of voice recognition to facilitate internet search on a mobile device, it does not provide technical details as to how the voice recognition works, such as “*generating a preliminary interpretation for the natural language utterance from the one or more syllables.*” Cross provides those details.

417. Specifically, Cross discloses that “ASR engine (150) is a module of computer program instructions, also stored in RAM in this example, that includes an ASR lexicon (106) of words capable of recognition by the ASR engine. The lexicon (106) is an association of words in text form with phonemes representing pronunciations of each word.” (Cross, 6:28-33.) As I discussed above, a POSITA would understand the reference to “pronunciations” to include syllables.

418. Cross further discloses “[t]he ASR engine [] uses the phonemes to find the word in the lexicon.” (Cross, 6:42-43.) The preliminary interpretation includes the recognized words or phrases. (Cross, 5:59-62 (“Multimodal server application (188) provides voice recognition services for multimodal devices by accepting requests for speech recognition and returning speech recognition results, including text representing recognized speech”).) Because the ASR lexicon is an “association of words in text form,” Cross discloses mapping phonemes to syllables and then generating a preliminary interpretation that includes a preliminary interpretation that includes words generated from the syllables.

419. The combination of Ramer and Cross discloses “*generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*”

## 6. Interpreting the Utterance Using Context [1E] / [27E]

[1E] interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

[27E] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

420. The combination of Ramer and Cross discloses “*interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance*” and “*a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

421. The ’176 patent provides describes what it considers “context” when “*interpreting the recognized words or phrases include[ing] establishing a context for the natural language utterance.*” The ’176 patent states that “as described in U.S.

patent application Ser. No. 11/200,164, an environmental model may be accessed to determine user location, user activity, track user actions, and/or other environmental information to invoke context, domain knowledge, preferences, and/or other cognitive qualities to enhance the interpretation of questions and/or commands.” (’176 patent, 4:64-5:3.)

422. The ’176 patent provides an example of the system using environmental information, like location, as context for interpreting the user’s natural language utterance: “Based on a context of the requested task (e.g., ‘Navigation,’ ‘Seattle,’ etc.), a voice search engine may retrieve an address of the Space Needle and a navigation application may calculate the route. Furthermore, user profile information may indicate that the user is visiting Seattle from out-of-town (e.g., the profile may indicate that the user’s home is Sacramento), and therefore, an advertisement for popular points-of-interest in Seattle may be selected.” (’176 patent, 8:55-63.)

423. Similar to the ’176 patent, Cross discloses interpreting the recognized words at a “*conversational language processor*.” The conversational language processor in Cross establishes a context for the utterance based on recent search history. (Cross, 13:57-62 (“context (304) is implemented as a data structure in which is stored words associated with user activity in the multimodal application, for

example, words uttered by a user as speech for recognition or words found on web sites previously visited by the user in operating the multimodal application”).)

424. Cross offers two examples where “*interpreting the recognized words or phrases includes establishing a context for the natural language utterance*”:

- a. For a user who spent the last few minutes browsing automobile web sites, for example, now opens a Google™ search window, a system that provides a context-based grammar for automatic speech recognition according to embodiments of the present invention effectively assumes that the search is related to automobiles and provides a voice recognition grammar that contains car names, makes, models, along with various other terms encountered in users browsing.

(Cross, 17:45-53.)

- b. As another example: For a user interacting with a system management console for a Z System mainframe who moves to a user group web site to do a search, the grammar for voice commands in that search box would be generated based on the text encountered in the system management console.

(Cross, 17:53-57.)

425. Ramer also discloses “*wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*” Like the ’176 patent, Ramer uses information about the user (e.g., location), as context to interpret the recognized words of the natural language utterance. As just one example, of the many disclosed in Ramer, the system uses location to establish context for the user’s utterance:

Similarly, disambiguation may proceed through facilities other than a disambiguation facility 140 (e.g., a parent controls facility 150 or algorithm facility 144). As shown in FIG. 3, 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.). This additional information, coupled with the original query entry 120, may permit an unambiguous query to be processed. For example, a user may enter a query entry 120 of “Royals.” This query entry 120 may be processed through the wireless communication facility 104 or a wireless provider 108 to a disambiguation facility 140 that is linked to a mobile subscriber database 112 containing, among other data, the user’s residence of Kansas City, Mo. The disambiguation facility 140 may link this demographic information to the query entry 120 “Royals” and 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. Disambiguation may include part-of-speech disambiguation, word sense disambiguation, phrase identification, named entry. recognition, or full sentential parsing. Part-of-speech disambiguation refers to the process of assigning a part-of-speech tag (e.g., noun, verb, adjective) to each word in a query. By assigning the part-of-speech tag to each word, the device can draw inferences about each word by virtue of its context. For example, the word “house” may be a noun or a verb. By tagging this word with an appropriate part-of-speech tag, additional information about the user query, and its ultimate goal, may be derived. Word sense disambiguation refers to the process of sorting words that have multiple meanings. Phrase identification refers to the process of relating each word to others within a phrase to derive the context of individual words. Named entity recognition generally refers to recognition of proper nouns that refer to specific names, places, countries, etc. Full sentential parsing is the process of decomposing a

sentence into smaller units and identifying the grammatical role of each and its relation to the other units. These and other techniques may be employed within the disambiguation facility 140 to infer a user's intended meaning for a search or search string.

(Ramer, 24:5-46.)

426. Ramer provides a similar example of using context about the user to interpret the recognized words and understand the implicit request the user is making even if it is not made explicitly by the user:

Sponsor information may be displayed to users based on the relation of the sponsor information and user search queries, results lists, items or categories, and the websites visited by the user. Web pages may display content for syndicated ads or links for syndicated ads. Furthermore, the wireless search platform 100 illustrated in FIG. 1 may contain the sorts of sponsor information described above in a separate database 128.

The implicit query facility 164 provides for the display of relevant content to users based on user activities other than explicit search queries. For example, in GPS data the locator facility 110 may indicate that the cell phone user is in the vicinity of a sponsor's restaurant. In addition, the clock contained in the mobile communication facility 102 and/or the wireless communication facility may indicate that it is mid evening. A predictive algorithm could merge this information and make the implicit query that the user is interested in restaurants in his immediate vicinity at which he could purchase dinner, and then push content (ads, phone numbers, menus, reviews) to his mobile communication facility 102 for immediate display.

(Ramer, 14:52-15:5.)

427. Ramer also discloses additional examples of establishing context in order to interpret the user's recognized words or phrases, including using the user's area code:

In one embodiment the search function is adapted based on area code. The area code-adapted search function may be an implicit query, an active query, a disambiguation action, a retrieval function, a filtering function, a presentation function, a routing function, or another function or action relating to the initiation, processing, or completion of a search. For example, a user with a given area code could receive implicit query results based on other searches by members of the same area code. Also, queries may be disambiguated or results filtered, sorted or presented based on area code (or other location information). For example, a user entering "Paris" in the 270 area code might receive results for Paris, Tenn.; a user entering "Paris" in the 310 area code might receive results for Paris, Hilton; and a user entering the same word in the 617 area code might receive results for Paris, France.

(Ramer, 27:63-28:10.)

428. The combination of Ramer and Cross thus discloses *"interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance."*

## **7. Selecting An Advertisement [1F] / [27F]**

[1F] selecting an advertisement in the context established for the natural language utterance; and

[27F] select an advertisement in the context established for the natural language utterance; and

429. The combination of Ramer and Cross discloses “*selecting an advertisement in the context established for the natural language utterance*” and a conversational language processor configured to “*select an advertisement in the context established for the natural language utterance.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

430. As discussed above, Ramer “relates to the field of mobile communications and more particularly to improved mobile advertisement syndication for mobile communication facilities.” (Ramer, 1:47-49.)

431. Ramer discloses “*selecting an advertisement in the context established for the natural language utterance*” by selecting advertisements based on information it knows about the user to provide context, such as the user’s past internet search history. (Ramer, 46:18-20 (“the bookstore in a building 1102B may want to advertise to the user 1104 because he is in the area and has some history of searching amazon.com and the like”).)

432. Ramer discloses using the context the system knows about the user to select targeted advertisements: “[S]earch results are presented to the user 1104 on the mobile communication facility 102 that are targeted to the user based on information relating to the user, the mobile communication facility 102, the location, and/or other information as described herein in conjunction with a personal filter. In embodiments, the information is also used to better target advertising, and sponsored

advertisements may be provided to the mobile communication facility 102 through a pay for auction advertisement scheme. Such a scheme could be similar to the desktop featured Google AdWords and AdSense by Google, Inc. of Mountain View, Calif.” (Ramer, 47:34-45.)

433. As Ramer discloses, the targeted advertising may be selected and presented to the user as part of the search results: “Once the initial explicit and/or implicit query is made, the query 212 is processed and the initial results retrieved 214. Both the query 212 and the initial retrieved results 214 may undergo additional filtering 258 and aggregation 260. Walled garden content 262 and sponsored content 220 may also attach to the query 212 and present tailored results 222 to the user. The results 222 may also, in turn, trigger the posting of additional sponsor 224 messages and advertisements.” (Ramer, 15:53-60.)

434. In addition, “[i]nformation from pay-per-click (PPC) sponsors 228 may link to the results, making it possible for the user to quickly learn about sponsors services, phone numbers, addresses, hours of operation, sales, and so forth.” (Ramer, 15:60-64.)

435. The combination of Ramer and Cross thus discloses “*selecting an advertisement in the context established for the natural language utterance.*”

## 8. Presenting the Selected Advertisement [1G] / [27G]

[1G] **presenting the selected advertisement via an output device** coupled to the conversational language processor.

[27G] **present the selected advertisement via an output device.**

436. The combination of Ramer and Cross discloses “*presenting the selected advertisement via an output device coupled to the conversational language processor*” and a conversational language processor configured to “*present the selected advertisement via an output device.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

437. Ramer discloses multiple times the advertisement may be presented: “The sponsored link may be presented upon retrieval of results, sorting of results, filtering of results, presentation of results, or routing of results, such as in response to a search function 142 that is executed in response to a query.” (Ramer, 48:22-25.)

438. Ramer further discloses that the system “[p]resents the selected advertisement via an output device”:

In embodiments, search results are presented to the user 1104 on the mobile communication facility 102 that are targeted to the user based on information relating to the user, the mobile communication facility 102, the location, and/or other information as described herein in conjunction with a personal filter. In embodiments, the information is also used to better target advertising, and sponsored advertisements may be provided to the mobile communication facility 102 through a pay for auction advertisement scheme. Such a scheme could be similar to

the desktop featured Google AdWords and AdSense by Google, Inc. of Mountain View, Calif.

(Ramer, 47:34-45.)

439. The combination of Ramer and Cross discloses “*presenting the selected advertisement via an output device coupled to the conversational language processor*” and a conversational language processor configured to “*present the selected advertisement via an output device.*”

### **C. Independent Claims 14 and 40**

440. Independent claims 14 and 40 include substantially overlapping processing limitations ([14A]-[14F]/[40A]-[40F]), as shown in the table below. Because of the overlap, I address these limitations together. Moreover, all but one (claim limitation [14F] and [40F]) of these limitations overlap with limitations addressed in independent claims 1 and 27, so those limitations have already been addressed above in §VIII.B. In this table, I also provide a citation to the section discussing each limitation.

| <b>Claim 1</b>                                                                                                                                                                                                                                               | <b>Claim 27</b>                                                                                                                                                                                                                                                                                                         | <b>Section</b> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| [14P] A method for selecting and presenting advertisements in response to processing natural language utterances, comprising:                                                                                                                                | [40P] A system for selecting and presenting advertisements in response to processing natural language utterances, comprising:                                                                                                                                                                                           | §VIII.C.1      |
| [14A] receiving a natural language utterance containing at least one request at an input device;                                                                                                                                                             | [40A] an input device that receives a natural language utterance containing at least one request at an input device;                                                                                                                                                                                                    | §VIII.C.2      |
| [14B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device;                                                                                                                 | [40B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance;                                                                                                                                         | §VIII.C.3      |
| [14C] interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | [40C] a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance; | §VIII.C.4      |
| [14D] selecting an advertisement in the context established for the natural language utterance;                                                                                                                                                              | [40D] select an advertisement in the context established for the natural language utterance; and                                                                                                                                                                                                                        | §VIII.C.5      |

| Claim 1                                                                                                                                                                                                                                                                                                             | Claim 27                                                                                                                                                                                                                                                                                                              | Section   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| [14E] presenting the selected advertisement via an output device coupled to the conversational language processor; and                                                                                                                                                                                              | [40E] present the selected advertisement via an output device; and                                                                                                                                                                                                                                                    | §VIII.C.6 |
| [14F] determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event. | [40F] an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event. | §VIII.C.7 |

#### 1. Preamble [14P] / [40P]

|                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>[14P] A method <b>for selecting and presenting advertisements in response to processing natural language utterances, comprising:</b></p> <p>[40P] A system <b>for selecting and presenting advertisements in response to processing natural language utterances, comprising:</b></p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

441. The combination of Ramer and Cross discloses “[a] *method for selecting and presenting advertisements in response to processing natural language utterances*” as well as a “*system for selecting and presenting advertisements in response to processing natural language utterances,*” as discussed below in §§VIII.C.1-7. The preambles for independent claims 14 and 40 are the exact same

as those for claims 1 and 27, respectively. For the same reasons discussed above in §VIII.B.1, the combination of Ramer and Cross discloses these limitations.

## 2. Receiving Natural Language Request [14A] / [40A]

[14A] receiving a natural language utterance containing at least one request at an input device;

[40A] an input device that receives a natural language utterance containing at least one request at an input device;

442. The combination of Ramer and Cross discloses “*receiving a natural language utterance containing at least one request at an input device*” as well as a “*an input device that receives a natural language utterance containing at least one request at an input device.*” Claim limitations [14A] and [40A] are the same as those for claim limitations [1A] and [27A], respectively. For the same reasons discussed above in §VIII.B.2, the combination of Ramer and Cross discloses these limitations.

## 3. Recognizing One or More Words in the Utterance [14B] / [40B]

[14B] recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device;

[40B] a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance;

443. The combination of Ramer and Cross discloses “*recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device*” as well as a “*a speech recognition engine*

*coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance.”* Claim limitations [14B] and [40B] are the same as those for claim limitations [1B] and [27B], respectively. For the same reasons discussed above in §VIII.B.3, the combination of Ramer and Cross discloses these limitations.

#### **4. Interpreting the Utterance Using Context [14C] / [40C]**

[14C] **interpreting the recognized words or phrases** at a **conversational language processor coupled to the speech recognition engine**, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

[40C] a **conversational language processor coupled to the speech recognition engine**, wherein the conversational language processor is configured to: **interpret the recognized words or phrases**, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance;

444. The combination of Ramer and Cross discloses “*interpreting the recognized words or phrases at a conversational language processor coupled to the speech recognition engine, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance*” as well as a “*a conversational language processor coupled to the speech recognition engine, wherein the conversational language processor is configured to: interpret the recognized words or phrases, wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*” Claim

limitations [14C] and [40C] are the same as those for claim limitations [1E] and [27E], respectively. For the same reasons discussed above in §VIII.B.6, the combination of Ramer and Cross discloses these limitations.

## **5. Selecting An Advertisement [14D] / [40D]**

[14D] selecting an advertisement in the context established for the natural language utterance;

[40D] select an advertisement in the context established for the natural language utterance; and

445. The combination of Ramer and Cross discloses “*selecting an advertisement in the context established for the natural language utterance*” as well as a conversational language processor configured to “*select an advertisement in the context established for the natural language utterance.*” Claim limitations [14D] and [40D] are the same as those for claim limitations [1F] and [27F], respectively. For the same reasons discussed above in §VIII.C.7, the combination of Ramer and Cross discloses these limitations.

## **6. Presenting the Selected Advertisement [14E] / [40E]**

[14E] presenting the selected advertisement via an output device coupled to the conversational language processor; and

[40E] present the selected advertisement via an output device; and

446. The combination of Ramer and Cross discloses “*presenting the selected advertisement via an output device coupled to the conversational language*

*processor” as well as a conversational language processor configured to “present the selected advertisement via an output device.” Claim limitations [14E] and [40E] are the same as those for claim limitations [1G] and [27G], respectively. For the same reasons discussed above in §VIII.B.8, the combination of Ramer and Cross discloses these limitations.*

**7. Determining An Incorrect Interpretation Occurred and Reinterpreting [14F] / [40F]**

[14F] determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

[40F] an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

447. The combination of Ramer and Cross discloses “*determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event*” as well as “*an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the*

*conversational language processor reinterprets the words or phrases in response to the predetermined event.” Both Ramer and Cross disclose these limitations.*

448. As noted above (§VIII.B.7) I understand from the ’176 patent that the inventors were aware that the user of an adaptive misrecognition engine was well known in the art and was not novel. A POSITA would therefore understand the inventors conceded that this limitation was already well known in the prior art.

449. In turn, Di Cristo notes that when the system misrecognizes a request, the user typically provides immediate feedback, such as overriding the command in a time shorter than the expected time execution time of the command, issuing a stop command, or taking other action. (Di Cristo, 2:60-66.) Di Cristo likewise describes various examples of responses that can be given by the misrecognition engine, including simply repeating the command back to the user:

When the system misrecognizes a request, the user typically provides immediate feedback such as, overriding the command in a time shorter than the expected time execution time of the command, repeating the original request, issuing a stop command, or taking other action, wherein the action may be presented verbally, non-verbally by pushing a selected button on a cell phone or remote control, or both, among other configurations.

(Di Cristo, 13:59-66.)

450. Ramer includes “disambiguation facility 140” which “may complete or provide more meaning to ambiguous active user inputs,” (Ramer, 9:14-15), as well as a correction process 244 which creates queries “sufficient to yield intelligible and

useful result set(s),” (Ramer, 24:49-53). The disambiguation and correction steps are shown in Figure 4 below.

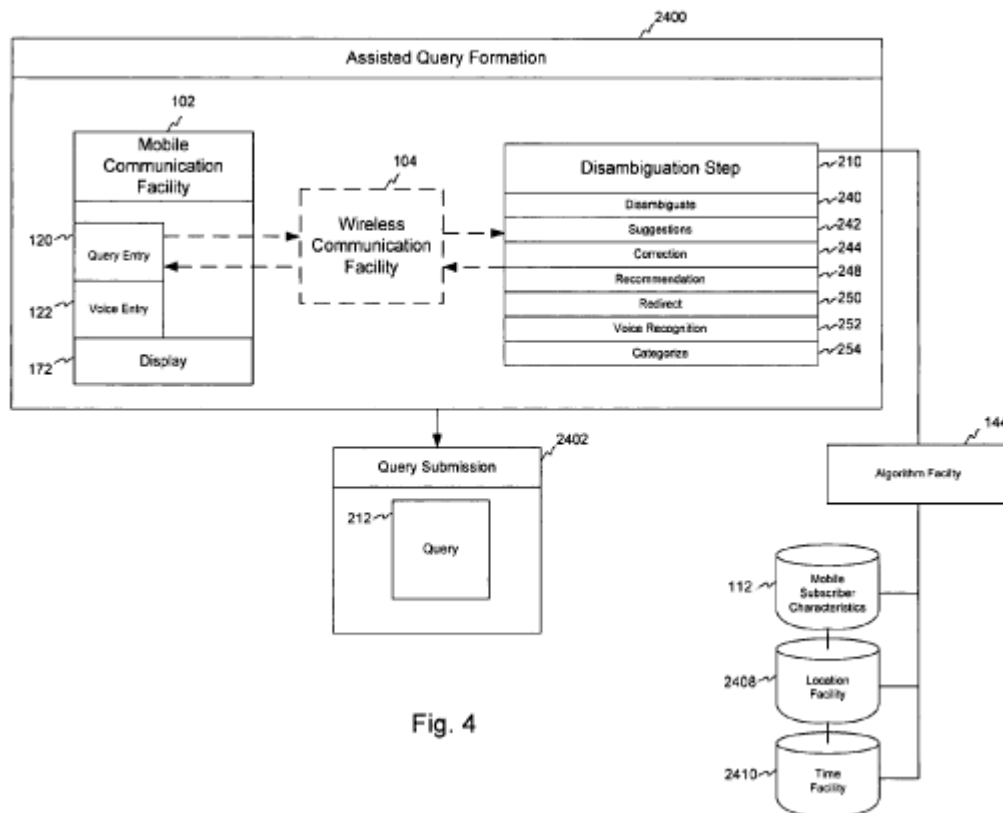


Fig. 4

**Ramer, Figure 4**

451. “Information stored in the mobile subscriber characteristics database 112, location information 2408, or time information 2410 may also be used with the correction 244 process.” (Ramer, 24:59-62.) Ramer thus discloses an “*adaptive misrecognition engine*” because the correction step can leverage database information that is being refreshed with more use.

452. Ramer further discloses that the “*adaptive misrecognition engine*” “*determines that the conversational language processor incorrectly interpreted the*

*words or phrases*” after “*detecting a predetermined event*” – namely that the user uttered a word that did not appear in its dataset.

453. In an illustrative example, a user utters “Coltrain” – “not a known word.” (Ramer, 25:9-15.) The system may misrecognize the word as “Coal Train” or “Soul Train.” (Ramer, 25:10-11.) The “query assistance facility 210 may predict a correction 244, taking into account user-specific data contained in the purchase history.” (Ramer, 25:11-13.) Specifically, “there may be a history of many recent purchases of compact discs recorded by the saxophonist, John Coltrane.” (Ramer, 25:7-9.) The system therefore, reinterprets the word from “Coltrain” to “Coltrane” because of the user’s purchase history. (Ramer, 25:3-19.) In doing so, the system “*reinterprets the words or phrases in response to the predetermined event.*”

454. In another example, Ramer discloses that, in those instances where “a user’s query entry 120 returns a null result set or an improbable result set,” “the search facility ... could automatically trigger correction 244 and iteratively cycle through alternative query entries 120 until a non-null or higher probability result set is delivered.” (Ramer, 25:20-25.)

455. Ramer provides additional details regarding how the misrecognition engine functions:

An interactive process between the mobile communication facility 102 and the disambiguation step 210 may be used for unambiguous query formation 2400. Once a user submits a query entry 120 to the mobile communication

facility 102, a process of correction 244 may be necessary or helpful for unambiguous query formation 2400 that is sufficient to yield intelligible and useful result set(s). As part of the correction 244 process, information specific to the type of mobile communication facility 102 may be used; for example, if the device has unique delivery capabilities, the query may need correction in order to derive a result set compatible with these capabilities. Information stored in the mobile subscriber characteristics database 112 or parental controls facility 150 may also be integral to the correction 244 process. In embodiments, the search engine may automatically suggest corrections for common misspellings, including those which are unique to the mobile experience (e.g., “2nite”).

A user’s query entry 120 may return a null result set or an improbable results set. In this case, the search facility 142, in conjunction with the mobile communication facility 102, could automatically trigger correction 244 and iteratively cycle through alternative query entries 120 until a non-null or higher probability result set is delivered.

(Ramer, 51:14-36.)

#### **D. Dependent Claims 2, 18, 28, and 44**

2. The method of claim 1, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

18. The method of claim 14, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a

specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

28. The system of claim 27, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

44. The system of claim 40, wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.

456. Dependent claims 2, 18, 28 and 44 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 2, 8, 28 and 44 add the following limitation: “*wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern*

*for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.”*

457. Ramer discloses “*wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific user.*” Specifically, Ramer discloses a “*conversational language processor selects the advertisement based on*” at least both (i) “*information related to one or more of the recognized words or phrases*” and (ii) “*an environmental model derived from environmental conditions or surroundings associated with the specific user.*”

458. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

459. As discussed above, Ramer discloses a system that interprets the user’s natural language request and uses the “*recognized words or phrases*” to select and

present an advertisement to the user. (*See supra* §§VIII.B.7-8). Ramer provides the following example where the recognized words or phrases are used to select the advertisement:

In embodiments, an advertisement may be presented to a mobile communication facility based at least in part on receiving a webpage request from the query facility of a mobile communication facility, receiving information associated with the mobile communication facility, and associating at least one advertisement with a webpage at least in part based on the information relating to the mobile communication facility. For example, a user of a mobile communication facility may initiate a search query for “Sonny Rollins.” The potential search results that may be presented to the user’s mobile communication facility may include advertisements and websites for CD’s of Sonny Rollin’s music, videos of his performances, etc. Information about the mobile communication facility (e.g., its video streaming capabilities) may be used in order to determine which of the advertisement/webpage results may be presented successfully to the user’s mobile communication facility. This information may, in turn, be used to pair webpages and advertisements that are each capable of presenting in the display of the user’s mobile communication facility.

(Ramer, 48:52-49:4.)

460. Also as discussed above, Ramer discloses the use of geographic information to select and present advertisements to users. (*See supra* §§ VIII.B.7-8.)

Ramer provides the following disclosure:

A user’s prior search activities and search results may also be used to create recommendations for the user. Prior search activities may include transactions, search queries, visits to websites, and other acts initiated by the user on

the mobile communication facility. **The geographic location of the mobile communication facility may foster recommendations including, but not limited to, sponsor information (e.g. products and services) in the user's current geographic vicinity.** The current time may be used independently or in conjunction with other information to create user recommendations. For example, the independent fact that it is noon, may create recommendations for restaurants serving lunch. This information may be further filtered by the location of the mobile communication facility to recommend only those restaurants that are in the user's immediate vicinity and further filtered by the subscriber's characteristics to recommend only that subset of restaurants serving lunch in the user's current vicinity that have received high ratings by restaurant patrons with a demographic profile similar to the user's. As with the above restaurant example, similar processes for generating meaningful recommendations may be applied to other services and products, including transportation, food, theater, sports, entertainment, movies, corporations, work, banks, post offices, mail facilities; location of and directions to gas stations, taxis, buses, trains, cars, airports, baby sitters, and other service and goods providers such as drug stores, drive through restaurants, bars, clubs; times of movies and entertainment; news; and information specific to the locale (directions, locations, starting times, new, etc.)

(Ramer, 51:49-52:11.)

461. The '176 patent describes an “*environmental model*” as follows: “For example, as described in U.S. patent application Ser. No. 11/200,164, an environmental model may be accessed to determine **user location**, user activity, track user actions, and/or other environmental information to invoke context, domain knowledge, preferences, and/or other cognitive qualities to enhance the

interpretation of questions and/or commands.” (’176 patent, 4:64-5:3.) As described in the example above, the system in Ramer determines a “*user location*” to enhance the interpretation of the user’s utterances. A POSITA would understand that the user location information and the time of day being accessed in the Ramer is part of an “*environmental model*” as described in the ’176 patent because it is an approximation or representation of an “aspect of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.”

462. Moreover, the ’176 patent admits that the use of an environmental model “to enhance the interpretation of questions and/or commands” was already known in the art and not a novel feature of the alleged invention. (’176 patent, 4:64-5:3.)

463. The combination of Ramer and Cross discloses “*wherein the conversational language processor selects the advertisement based on information related to one or more of the recognized words or phrases, an action associated with the request, a personalized cognitive model derived from an interaction pattern for a specific user, a generalized cognitive model derived from an interaction pattern for a plurality of users, or an environmental model derived from environmental conditions or surroundings associated with the specific use.*”

#### **E. Dependent Claims 3, 19, 29 and 45**

## 1. Preamble [3P], [19P], [29P] and [45P]

3. The method of claim 2, further comprising:

19. The method of claim 18, further comprising:

29. The system of claim 28, wherein the conversational language processor is further configured to:

45. The system of claim 44, wherein the conversational language processor is further configured to:

464. The preambles for claims 3, 19, 29 and 45 simply refer back to the claims from which they depend and do not add any substantive limitations. As discussed above in §VIII.D regarding claims 2, 18, 28 and 44, those claims are all disclosed in the prior art and would have been well known to a POSITA.

## 2. Tracking an Interaction Pattern [3A], [19A], [29A] and [45A]

[3A] tracking an interaction pattern with the advertisement presented via the output device; and

[19A] tracking an interaction pattern with the advertisement presented via the output device; and

[29A] track an interaction pattern with the advertisement presented via the output device; and

[45A] track an interaction pattern with the advertisement presented via the output device; and

465. The combination of Ramer and Cross discloses “*tracking an interaction pattern with the advertisement presented via the output device*” as well as conversational language processor is configured to “*track an interaction pattern with*

*the advertisement presented via the output device.”* Specifically, Ramer discloses these limitations.

466. Ramer discloses that the system tracks users interactions with information provided to users, including “*tracking an interaction pattern with an advertisement*”:

In embodiments, a basic implementation of a collaborative filtering algorithm entails **tracking the popularity of a product, service, business, transaction, or website by recording the total number of users in the set that rate it favorably** (which may include a degree of favorability) or by recording the number of users that repeatedly visit the site. In essence, this algorithm assumes that what previous users liked, new users will like. In embodiments, a weighted averaging process is implemented to carve out Subgroups of users who all highly ranked a product that has an overall unpopular rating with the overall user dataset.

(Ramer, 11:1-11.)

467. Ramer provides additional disclosure how the system tracks interaction patterns with advertisements, and explains that that information is stored in a database:

In embodiments, **advertising and webpage content compatibility with a type of mobile communication facility 102 may be determined at least in part by tracking a plurality of mobile communication facility 102 interactions with such content.** Information may be stored pertaining to the advertising and web interactions in a database, where a portion of the information comprises identification of at least one mobile communication facility 102 from the plurality of mobile communication

facilities, and predicting the compatibility of the mobile communication facility 102 with the content based, in part, on how many content interactions there were. The prediction of compatibility of the mobile communication facility 102 with content may be based, in part, on how many interactions there were in the user's past. Content may be a download, program, file, executable file, Zipped file, compressed file, audio, and video. An advertising or web interaction may be a click on a hyperlink, an indication of downloaded content, and/or an indication of a downloaded program.

(Ramer, 142:10-28.)

468. Ramer further discloses how user tracking is used to optimize search results, which include advertisements:

In embodiments, a method of optimizing search results for mobile users may include tracking the on-line interactions of a mobile communication facility 102, where the tracking involves, in part, identifying an identification number associated with the mobile communication facility 102, generating a filter based, in part, on the tracked on-line interactions, and applying the filter to a search performed on the mobile communication facility 102. The identification number may include a phone number. The tracked on-line information may include tracking clicks, clickthroughs, queries, clicks following queries, WAP sites visited, WAP portals visited, information reviewed from a DEC directory (e.g., a carrier's catalog), information reviewed from a billing history associated with a user, information about payment methods, purchases, payment timing, timing of online interactions and/or the location, and speed and direction of the mobile communication facility 102 at the time of the online interaction.

(Ramer, 124:5-21.)

**3. Updating a Model Based on the Interaction Pattern [3B], [19B], [29B] and [45B]**

[3B] updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

[19B] updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

[29B] update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

[45B] update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.

469. The combination of Ramer and Cross “*updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement*” as well as conversational language processor is configured to “*update one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model based on the interaction pattern tracked for the advertisement.*” Specifically, Ramer discloses these limitations.

470. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other

characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

471. After the system described in Ramer collects the “*interaction pattern tracked for the advertisement*” it stores the information and uses it or future interactions with users. The information is stored in a database that includes the models which the system uses to select an advertisement, including “location, personal information relating to a user, web interactions, email interactions, messaging interactions, billing history, payment history, typical bill amount, time of day, [etc.]” (Ramer, 50:25-35.) Ramer is at least describing an “*environmental model based on the interaction pattern tracked for the advertisement*” because the system is updating the environmental model to provide the user’s location and time of day. (Ramer, 50:25-35.)

472. Ramer also teaches that its algorithm is designed to assume “that what previous users liked, new users will like.” (Ramer, 11:1-11.) For Ramer to use previous user’s likes to predict future user’s likes it must necessarily update its model. Ramer discloses Ramer also discloses that “[i]nformation may be stored pertaining to the advertising and web interactions in a database,” (Ramer, 142:10-29), and that the that the statistical models are continually updated:

In embodiments, mobile communication facility users may be classified on a combination of mobile subscriber characteristics, device type, location, behavioral history, transaction history, or other parameters and the resulting

classes presented in a menu format to sponsors for bidding. For example, it may be possible to statistically model the user characteristics within a population of a wireless provider's customer base that are most likely to purchase tickets for a Caribbean cruise. Once this model is known, all customers with approximately the profile described by the model could be grouped in a "Caribbean Cruisers" category. This category may then be included in a menu system, along with other categories, and presented to sponsors for bidding. Swimwear companies, sunglass companies, etc. may have an increased interest in bidding for access to the Caribbean Cruisers category because of the increased probability that users in this category have a need for their products. Summary of the many characteristics (variables) used in the model into a single category may make the bidding process less time-consuming to sponsors and may constitute a proprietary product. **Furthermore, the statistical model may be continually updated to accommodate changing user preferences.**

(Ramer, 139:7-28.)

473. By updating the database with user information the system is "*updating ... the environmental model based on the interaction pattern tracked for the advertisement*" because user interactions with advertisements are tracked and used by the system to select and present advertisements in the future. For example, the system in Ramer may associate keywords with advertisements that are tied to a specific geographic area or times of day. (Ramer, 51:49-52:11.)

#### **F. Dependent Claims 4, 20, 30 and 46**

|                                                                                                                                                                                          |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [4] The method of claim 3, wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

selecting subsequent advertisements in response to subsequent natural language utterances.

[20] The method of claim 19, wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

[30] The system of claim 29, wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

[46] The system of claim 45, wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.

474. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

475. The combination of Ramer and Cross discloses “*wherein updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model builds statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances*” as well as

*“wherein the conversational language processor updates one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.”* Specifically, Ramer discloses these limitations.

476. Ramer discloses details regarding how the statistical profiles are built in the system:

A more robust collaborate filtering procedure, sometimes referred to as the K-nearest neighbor algorithm, uses a ‘training data set’ that is based upon previous users’ behavior to predict a variable of interest to members of a ‘target data set’ comprised of new users. In addition to user preference data, the training data set may have additional predictor variables, such as might be contained in a mobile subscriber characteristics database (e.g., age, income, sex, date or place of birth, etc.). Variables of interest may include type of product purchased, amount of purchase, and so forth. For each row (single user data) in the target data set, the algorithm locates the “K” closest members of the training data set. Closeness, or distance, as used by the algorithm is generally a Euclidean Distance measure. Next, the algorithm finds the weighted sum of the variable of interest for the K nearest neighbors, where the weights are the inverse of the calculated distances. This process is then repeated for all remaining rows in the target set. From this information, models may be derived for future prediction. As the user population increases, the training data set may be updated to include new rows and thus capture any changes in user preference for use in revising the prediction model.

(Ramer, 11:12-33.)

477. Ramer also discloses collecting preference data from a “large group of users” to be “analyzed statistically to identify subgroups, or characteristics of subgroup members, with similar preference profiles.” (Ramer, 10:14-18.) Ramer teaches that “[v]arious weighted average, fuzzy logic, or other techniques may be used to summarize or model” the data. (Ramer, 10:18-21.)

478. Ramer further discloses: “A recommendation system may use information from a user’s profile to make predictions regarding other information/products that might interest the user.” (Ramer, 12:11-13.)

#### **G. Dependent Claims 5, 21, 31 and 47**

[5] The method of claim 4, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

[21] The method of claim 20, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

[31] The system of claim 30, wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.

[47] The system of claim 46, wherein the statistical profiles identify affinities between the advertisement presented via the

|                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model. |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

479. The combination of Ramer and Cross discloses “*wherein the statistical profiles identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.*” Specifically, Ramer discloses this limitation.

480. As discussed above, I understand that in the Amazon Litigation the parties agreed that the term “model” means “an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system.” I have applied that construction in my analysis of these claims.

481. As I discussed above (§VIII.A.1), the Ramer system is designed to assume “that what previous users liked, new users will like,” which necessarily requires the system to identify “*affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.*” (Ramer, 11:1-11.) Ramer also discloses that the data “obtained from [] similar users provide the empiric basis for predicting information of relevance to the active user,” which demonstrates Ramer

identifying affinities between advertisements and “*the generalized cognitive model.*”  
(Ramer, 10:14-23.)

482. Ramer discloses that affinities are identified between advertisements and the actions associated with the models: “Finally, the advertiser may wish that the ad(s) target only a subset of users, such as, only users residing in a particular geographic region, meeting a set of demographic user characteristics, and/or users with a particular mobile communication facility 102.” (Ramer, 61:23-27.)

483. Ramer further discloses:

A user’s prior search activities and search results may also be used to create recommendations for the user. Prior search activities may include transactions, search queries, visits to websites, phone calls, and/or other acts initiated by the user in the mobile communication facility 102. The geographic location of the mobile communication facility 102 may foster recommendations including, but not limited to, sponsor information (e.g. products and services) in the user’s current geographic vicinity. The current time may be used independently or in conjunction with other information to create user recommendations. For example, the independent fact that it is noon may create recommendations for restaurants serving lunch. This information may be further filtered by the location of the mobile communication facility 102 to recommend only those restaurants that are in the user’s immediate vicinity, and it may be further filtered by the subscriber’s characteristics to recommend only that Subset of restaurants serving lunch in the user’s current vicinity that have received high ratings by restaurant patrons with a demographic profile similar to the user’s. As with the above restaurant example, similar processes for generating meaningful recommendations may be applied to other

services and products, including transportation (navigation, taxis, buses, trains, cars, airports, etc.), food and drink (groceries, drive through restaurants, bars, etc.), entertainment (theater, sports, movies, clubs, etc.), business (corporations, workplaces, banks, post offices and other mailing or shipping facilities. etc.), consumer needs (gas stations, drug or clothing stores, baby sitters, parking, etc.), and information specific to the locale (directions, locations, starting times, news, etc.)

(Ramer, 51:49-52:11.)

484. Ramer also discloses that various information regarding users can be collected and used to generate search results and for advertising: “In embodiments, the information is also used to better target advertising, and sponsored advertisements may be provided to the mobile communication facility 102 through a pay for auction advertisement scheme. Such a scheme could be similar to the desktop featured Google AdWords and AdSense by Google, Inc. of Mountain View, Calif.” (Ramer, 47:39-45.)

485. Ramer discloses that various types of information that indicate *“affinities between the advertisement presented via the output device and one or more of the recognized words or phrases, the action associated with the request, the personalized cognitive model, the generalized cognitive model, or the environmental model.”* For example, Ramer discloses:

A user’s prior search activities and search results may also be used to create recommendations for the user. Prior search activities may include transactions, search queries, visits to websites, and other acts initiated by the user on

the mobile communication facility. The geographic location of the mobile communication facility may foster recommendations including, but not limited to, sponsor information (e.g. products and services) in the user's current geographic vicinity. The current time may be used independently or in conjunction with other information to create user recommendations. For example, the independent fact that it is noon, may create recommendations for restaurants serving lunch. This information may be further filtered by the location of the mobile communication facility to recommend only those restaurants that are in the user's immediate vicinity and further filtered by the subscriber's characteristics to recommend only that subset of restaurants serving lunch in the user's current vicinity that have received high ratings by restaurant patrons with a demographic profile similar to the users. As with the above restaurant example, similar processes for generating meaningful recommendations may be applied to other services and products, including transportation, food, theater, sports, entertainment, movies, corporations, work, banks, post offices, mail facilities; location of and directions to gas stations, taxis, buses, trains, cars, airports, babysitters, and other service and goods providers such as drug stores, drive through restaurants, bars, clubs; times of movies and entertainment; news; and local information.

(Ramer, 25:35-62.)

486. Ramer similarly discloses:

Sponsor information may be displayed to users based on the relation of the sponsor information and user search queries, results lists, items or categories, and the websites visited by the user. Web pages may display content for syndicated ads or links for syndicated ads. Furthermore, the wireless search platform 100 illustrated in FIG. 1 may contain the sorts of sponsor information described above in a separate database 128. The implicit query facility 164 provides for the display of relevant content to users based

on user activities other than explicit search queries. For example, in GPS data the locator facility 110 may indicate that the cell phone user is in the vicinity of a sponsor's restaurant. In addition, the clock contained in the mobile communication facility 102 and/or the wireless communication facility may indicate that it is mid evening. A predictive algorithm could merge this information and make the implicit query that the user is interested in restaurants in his immediate vicinity at which he could purchase dinner, and then push content (ads, phone numbers, menus, reviews) to his mobile communication facility 102 for immediate display.

(Ramer, 14:52-15:5.)

487. Ramer also discloses adapting its search function based on: age (column 26), gender (column 27), ethnicity (column 27), area code (column 27), home address (column 28), work address (column 28), billing address (column 28), current location (column 28), credit card info (column 29), family information (column 29), employment data (column 30), employer (column 30), annual income (column 30), items purchased (column 31), friends/family (column 31), average bill/payment history/on-time payment (column 32), on-line usage (column 32), duration of on-line interaction (column 32), shopping habits (column 34), click stream info (column 34), history of click throughs (column 42).

## **H. Dependent Claims 6, 22, 32 and 48**

### **1. Preamble [6P], [22P], [32P] and [48P]**

[6P] The method of claim 3, further comprising:

[22P] The method of claim 19, further comprising:

[32P] The system of claim 29, wherein the conversational language processor is further configured to:

[48P] The system of claim 45, wherein the conversational language processor is further configured to:

488. The preambles for claims 6, 22, 32 and 48 simply refer back to the claims from which they depend and do not add any substantive limitations. As discussed above in §VIII.E regarding claims 3, 19, 29 and 45, those claims are all disclosed in the prior art and would have been well known to a POSITA.

### **2. Building Long-Term and Short-Term Knowledge [6A], [22A], [32A] and [48A]**

[6A] building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model;

[22A] building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model;

[32A] build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model; and

|                                                                                                                                                                                                                  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [48A] build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model; and |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

489. The combination of Ramer and Cross discloses “*building long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model*” as well as a conversational language processor configured to “*build long-term shared knowledge and short-term shared knowledge in response to updating one or more of the personalized cognitive model, the generalized cognitive model, or the environmental model.*” Specifically, Ramer discloses these limitations.

490. Ramer discloses a system that can “*build long-term shared knowledge and short-term shared knowledge*” in response to updating each of a personalized cognitive model, a generalized cognitive model, and an environmental model.

491. For example, Ramer discloses updating the personalized cognitive model as it learns the speech patterns of the user:

A voice recognition facility 160 may be a software component enabling a machine or device (e.g., a cellular phone) to understand human spoken language and to carry out spoken commands. Typically, a human voice is received by the device and converted to analog audio. The analog audio may in turn be converted into a digital format using, for example, an analog-to-digital converter, which digital data may be interpreted using voice recognition techniques. Generally this is done 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. The speech patterns of a unique user may be stored on a hard drive (locally or remotely) or other memory device, and may be loaded into memory, in whole or in part, when the program is run. A comparator may use, for example, correlation or other discrete Fourier transform or statistical techniques to compare the stored patterns against the output of the analog-digital converter.

(Ramer, 21:22-39.)

492. A POSITA would understand that the system builds long-term knowledge about the user's speech pattern and short-term knowledge related to the user's current request in response to updating the personalized cognitive model for the user. Ramer also provides an example of building long term and short term knowledge about the user:

For example, a user may provide a query entry 120 "Coltrain" that is sent through a wireless communication facility 104 so that the query assistance facility 210 may begin. Because the query assistance facility 210 is linked to other data sources, as part of the correction 244 process, information from the filter algorithm facility 144 may be linked to the query entry 120 to provide more information to be used in the process of producing a more useful search query. In this example, the user's filter algorithm facility 144 may use information such as a history of online purchases, product names, numbers, purchase amounts, and purchase dates and times. Within the databases associated with the filter algorithm facility 144 there may be a history of many recent purchases of compact discs recorded by the saxophonist, John Coltrane. Because the original user query entry 120, "Coltrain" is not a known word, the query assistance facility 210 may predict a correction 244, taking into account user-specific data contained in the purchase history of the filter algorithm

facility 144, such as that the user seeks information pertaining to “Coltrane,” as opposed to “Coal Train” or “Soul Train,” etc. In various embodiments, or particular user searches within one embodiment, the query assistance facility 210 may involve one of these additional data facilities, a plurality of the data facilities, or none of the data facilities.

(Ramer, 24:63-25:19)

493. Ramer also discloses a system that can “*build long-term shared knowledge and short-term shared knowledge*” in response to updating a generalized cognitive model, which provides a statistical abstract about general user behavior. As discussed above, Ramer builds long term knowledge in response to learning information about general user behavior to assist the system in disambiguating user utterances. As discussed above (§VIII.A.1), Ramer also discloses building long term information regarding users in response to updating the generalized cognitive model as discussed above in claim 5. As discussed above, Ramer updates the generalized cognitive model to include information regarding the following user information: age (column 26), gender (column 27), ethnicity (column 27), area code (column 27), home address (column 28), work address (column 28), billing address (column 28), current location (column 28), credit card info (column 29), family information (column 29), employment data (column 30), employer (column 30), annual income (column 30), items purchased (column 31), friends/family (column 31), average bill/payment history/on-time payment (column 32), on-line usage (column 32),

duration of on-line interaction (column 32), shopping habits (column 34), click stream info (column 34), history of click throughs (column 42). (*See also* Ramer, 26:45-27:27.)

494. Ramer also discloses a system that can “*build long-term shared knowledge and short-term shared knowledge*” in response to updating an environmental model. Also as discussed above, Ramer discloses the building long term information about user’s geographic location and time (environmental information). (*See supra* §§ VIII.B.7-8.) Ramer provides the following disclosure:

A user’s prior search activities and search results may also be used to create recommendations for the user. Prior search activities may include transactions, search queries, visits to websites, and other acts initiated by the user on the mobile communication facility. **The geographic location of the mobile communication facility may foster recommendations including, but not limited to, sponsor information (e.g. products and services) in the user’s current geographic vicinity. The current time may be used independently or in conjunction with other information to create user recommendations.** For example, the independent fact that it is noon, may create recommendations for restaurants serving lunch. This information may be further filtered by the location of the mobile communication facility to recommend only those restaurants that are in the user’s immediate vicinity and further filtered by the subscriber’s characteristics to recommend only that subset of restaurants serving lunch in the user’s current vicinity that have received high ratings by restaurant patrons with a demographic profile similar to the user’s. As with the above restaurant example, similar processes for generating meaningful recommendations may be applied to other services and products, including transportation, food, theater, sports,

entertainment, movies, corporations, work, banks, post offices, mail facilities; location of and directions to gas stations, taxis, buses, trains, cars, airports, baby sitters, and other service and goods providers such as drug stores, drive through restaurants, bars, clubs; times of movies and entertainment; news; and local information.

(Ramer, 25:35-62.)

### 3. Receiving a Subsequent Natural Language Utterance [6B] and [22B]

[6B] receiving a subsequent natural language utterance at the input device; and

[22B] receiving a subsequent natural language utterance at the input device; and

495. The combination of Ramer and Cross discloses “*receiving a subsequent natural language utterance at the input device.*” Specifically, Ramer discloses this limitation.

496. In the example described above in §VIII.A.1, Ramer discloses building long term knowledge about a user’s speech pattern to enhance the system’s interpretation of user requests. As the system disclosed in Ramer is designed to learn about the user’s behavior for future interactions with the user, it necessarily is “*receiving a subsequent natural language utterance at the input device.*”

An age-adapted search function may also operate in connection with an active query; for example, a given query may return results that are age-appropriate, age-filtered, age-ranked, or age-disambiguated. For example, entering “Pink” might return results for Pink Floyd if the user is over age 30, while it might return results for the

female artist Pink if the user is under 30. A user might be **prompted to resolve such an ambiguity**, or the ambiguity might be automatically resolved for the user.

(Ramer, 26:62-27:3.) A POSITA would understand that the user presents the “*subsequent natural language utterance*” at the “*input device*” (e.g., mobile phone).

#### 4. Interpreting Using Long-Term and Short-Term Knowledge [6C], [22C], [32B] and [48B]

[6C] interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.

[22C] interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.

[32B] interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.

[48B] interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.

497. The combination of Ramer and Cross discloses “*interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge,*” as well as a conversational language processor configured to “*interpret a subsequent natural language utterance received at the input device language processor using the long-term shared knowledge and the short-term shared knowledge.*” Specifically, Ramer discloses this limitation.

498. As discussed above, Ramer provides numerous examples of *“interpreting the subsequent natural language utterance at the conversational language processor using the long-term shared knowledge and the short-term shared knowledge.”* As just one example, as discussed above, Ramer discloses that the system interprets the subsequent utterance using long term shared knowledge (e.g., “age-adapted search function”) and short term shared knowledge (user request for about “Pink”). (Ramer, 26:62-27:7.) Likewise, the system interprets a user request of “Coltrain” by using short term knowledge about the current conversation and long-term knowledge about the user’s past behavior to interpret the user’s request:

For example, a user may provide a query entry 120 “Coltrain” that is sent through a wireless communication facility 104 so that the query assistance facility 210 may begin. Because the query assistance facility 210 is linked to other data sources, as part of the correction 244 process, information from the filter algorithm facility 144 may be linked to the query entry 120 to provide more information to be used in the process of producing a more useful search query. In this example, the user’s filter algorithm facility 144 may use information such as a history of online purchases, product names, numbers, purchase amounts, and purchase dates and times. Within the databases associated with the filter algorithm facility 144 there may be a history of many recent purchases of compact discs recorded by the saxophonist, John Coltrane. Because the original user query entry 120, “Coltrain” is not a known word, the query assistance facility 210 may predict a correction 244, taking into account user-specific data contained in the purchase history of the filter algorithm facility 144, such as that the user seeks information

pertaining to “Coltrane,” as opposed to “Coal Train” or “Soul Train,” etc. In various embodiments, or particular user searches within one embodiment, the query assistance facility 210 may involve one of these additional data facilities, a plurality of the data facilities, or none of the data facilities.

(Ramer, 24:63-25:19.)

#### **I. Dependent Claims 7, 23, 33 and 49**

[7] The method of claim 3, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

[23] The method of claim 19, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

[33] The system of claim 29, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

[49] The system of claim 45, wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.

499. The combination of Ramer and Cross discloses “*wherein the interaction pattern tracked for the advertisement includes an action performed in response to a subsequent request that identifies the advertisement.*” Specifically, Ramer discloses this limitation.

500. As discussed above in §VIII.E, Ramer discloses an “*tracking an interaction pattern with the advertisement,*” which is claimed in claim 3. Ramer also discloses that “*the interaction pattern tracked for the advertisements*” includes “*an action performed in response to a subsequent request that identifies the advertisement.*” Specifically, Ramer discloses that the search function, including providing advertisements in the search results is based on items purchased, which is an “*action performed in response to a subsequent request that identifies the advertisement*”:

In one embodiment the search function is adapted based on items purchased. The items-purchased-adapted search function may be an implicit query, an active query, a disambiguation action, a retrieval function, a filtering function, a presentation function, a routing function, or another function or action relating to the initiation, processing, or completion of a search. Items purchased may be obtained from a database maintained by a carrier, such as showing transactions made using a mobile communication facility 102. For example, implicit queries may be run and results presented based on transaction history, such as presenting results for peripherals and software applications for individuals who have recently purchased a computer, or the like. Similarly, queries may be disambiguated, or search results retrieved, sorted, presented, or routed based on items purchased by a user. For example, 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.

(Ramer, 31:23-42)

501. Ramer further discloses tracking user “shopping habits,” a similar “action performed in response to a subsequent request that identifies the advertisement” and using that information to adapt search results, including advertisements:

In one embodiment the search function 142 is adapted based on shopping habits (e.g., views of or purchases of goods and services made with a technology like Mobile Lime). The shopping-habit-adapted search function 142 may be an implicit query 164, an active query, a disambiguation action, a retrieval function, a filtering function, a presentation function, a routing function, or another function or action relating to the initiation, processing, or completion of a search. For example, a user who shops frequently for music may be presented with music-oriented content in an implicit search 204, while the implicit search may present different items to other users whose shopping habits are different. Similarly, queries may be disambiguated, or results returned, sorted, presented, or routed, based on shopping habits. For example, a frequent book purchaser may be presented with Stephen King books upon entering “King” in a query interface, while a frequent music purchaser may be presented with results related to Elvis Presley.

(Ramer, 34:4-21.)

502. Ramer provides additional information regarding how user’s interaction pattern is tracked for advertisements:

User preferences may be derived from user behavior or other implicit characteristics, or explicitly defined by a mobile communication facility user, or some combination of these. If users were to explicitly state their preferences (e.g. for types of restaurants, books, e-commerce, music, news, video, formats, audio, etc.), the explicit preference

information may be stored in the mobile subscriber characteristic data bases associated with their phones. Users may implicitly register a preference through activity such as purchasing a product online, visiting a site on line, making a phone call from a mobile communication facility, making a phone call from another facility, viewing content, or engaging or not engaging in other activities. For example, if the user looks at a product and decides not to purchase the product, one can draw an inference that the user is not interested in the product, and this inference may be used as part of a collaborative filtering algorithm. In addition, inferences may be drawn from the types of establishments the user has been calling recently on the mobile communication facility. If he or she has been calling auto dealerships repeatedly over the past two weeks, an inference can be drawn that the user is presently looking for auto goods and/or services. Implicit preferences of users may also be collected by recording all pages that are visited by users and the frequency and/or duration of each visit. Using a binary coding scheme in which visited pages are coded "1" and unvisited pages "0," one may create user-based preference vectors and analyze statistically for both intra-user and inter-user cluster preferences or similarities. Other coding techniques may group certain sites along dimensions of commonality, with navigation behavior analyzed using any number of Euclidean or other distance and/or matching techniques. In embodiments, user preference data may be collected from within the mobile subscriber characteristics database. In embodiments, user preference data may be collected from outside of the mobile subscriber characteristics database. In embodiments, off-line behavior may also be used to characterize the preferences of the user.

(Ramer, 10:31-67.)

## J. Dependent Claims 8, 24, 34 and 50

[8] The method of claim 7, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

[24] The method of claim 23, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

[34] The system of claim 33, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

[50] The system of claim 49, wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.

503. The combination of Ramer and Cross discloses “*wherein the action includes executing a task or retrieving information based on the subsequent request that identifies the advertisement.*” Specifically, Ramer discloses this limitation.

504. As discussed above in §VIII.E, Ramer teaches that the system tracks the “interaction pattern for the advertisement” including “*executing a task or retrieving information based on the subsequent request that identifies the advertisement,*” by providing the user an opportunity to purchase items through a link after the user specifically identifies the advertisement:

In embodiments of the present invention, methods and systems may include presenting a link in a user interface of a mobile communication device, where the link is configured to link to a commercial item that is available via a computer network, thus enabling a party to sponsor the link. The link may be to an item of goods or services for sale, to a promotion, to a content item, to an

advertisement, or to other material of a commercial provider, such as a vendor of goods or services. The link may, for example, be to an item on an electronic commerce site, to an auction site, to a reverse auction site, to a news site, to an information site, or to other content on a computer network, such as web content, content located on other networks, or the like. The link may include or be associated with various e-commerce features, such as those enabling single click purchasing, bidding, targeted advertising, instant purchasing (e.g., “buy it now”), tracking of clicks or transactions, tracking of referrals, affiliate program features, or the like. In embodiments the link is a sponsored link that is presented to a user on the user interface of a mobile communication facility 102, such as a cell phone. The sponsored link may be presented in association with an implicit query 164 (and may be related to such a query). **The sponsored link may be presented in response to entering an explicit query (or partial entry of such a query).** The sponsored link may be presented upon making of a telephone call or other action of a user of the mobile communication facility 102. The sponsored link may be presented upon retrieval of results, sorting of results, filtering of results, presentation of results, or routing of results, such as in response to a search function 142 that is executed in response to a query.

(Ramer, 47:63-48:25.)

505. A POSITA would understand that “an explicit query” means that the user explicitly requested the information related to the sponsored link.

506. Also as discussed above, Ramer discloses that clicks and click-throughs are tracked by the system. (Ramer, 124:13-15.) Therefore, the system disclosed in Ramer tracks an interaction pattern for the advertisement, including “*executing a*

*task or retrieving information based on the subsequent request that identifies the advertisement.”*

**K. Dependent Claims 9, 17, 35 and 43**

9. The method of claim 1, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

17. The method of claim 14, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

35. The system of claim 27, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

43. The system of claim 40, wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.

507. Dependent claims 9, 17, 35 and 43 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 9, 17, 35 and 43 add the following limitation: “*wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.*”

508. As noted above (§VIII.C.7), Ramer discloses that the system can resolve the user’s request “*in response to determining that the natural language*

*utterance includes incomplete or ambiguous information.*” For example, Ramer resolves ambiguity in whether a user is seeing information about “Pink Floyd” or the solo artist “Pink” in response to a query concerning music from “Pink.” (Ramer, 26:45-27:7.)

509. A POSITA would understand that Ramer discloses that targeted advertisements would also be selected and presented along with the search results as discussed above, so Ramer discloses that the system “*selects the advertisement to resolve the request.*” (Ramer, 140, 6-9 (“The potential search results that may be presented to the user’s mobile communication facility 102 may include advertisements 2002a, 2002b and websites for the musician’s CDs, videos of his performances, etc.”).)

510. The combination of Ramer and Cross discloses “*wherein the conversational language processor selects the advertisement to resolve the request in response to determining that the natural language utterance includes incomplete or ambiguous information.*”

## **L. Dependent Claims 10, 25, 36 and 51**

10. The method of claim 1, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

25. The method of claim 14, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

36. The system of claim 27 wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

51. The system of claim 40, wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.

511. Dependent claims 10, 25, 36 and 51 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 10, 25, 36 and 51 add the following limitation: “*wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary*

*and phrase tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.”*

512. Cross discloses recognizing words in the utterance “*using a plurality of entries in one or more dictionary and phrase tables.*” In one embodiment,

[C]reating (302) a context (304) is carried out by ... generating (326) a list (328) of most recently used words in the history, and selecting (330) words for the context (304) from the history in dependence upon the list of most recently used words. Generating (326) a list of most recently used words in the history may be implemented by storing words from previously visited web pages in association with time stamps indicating the **most recent time when a user uttered the word as speech for recognition** and the word was recognized by the ASR engine (150).

513. (Cross, 15:27-41.) The “*dictionary and phrase tables*” are “dynamically updated based on a history of a current dialog and one or more prior dialogs.” (Cross, 15:23-26 (“the multimodal application (189) adds words from the context (304) to the grammar (104), instructs an ASR engine (150) to compile the grammar into a lexicon (106), and so on”).)

514. Ramer similarly discloses the use of a “user specific dictionary” that is generated to assist the system in interpreting the user’s utterances:

In embodiments, text may be predicted and associated with a search query entered on a mobile communication facility 102 using voice recognition. The predictive text may be based, in part, on at least one of mobile communication facility 102 information, SMS conversion, mobile subscriber characteristic information, **user specific**

**dictionary**, numeric keypad conversion, and multi-letter keypad conversion.

...

In embodiments, text may be predicted and associated with a search query entered on a mobile communication facility 102, where the predicted text may be based, in part, on a search query entered through a voice recognition facility 160 residing on the mobile communication facility 102. Alternatively, the voice recognition facility 160 may reside on a remote server or in part on the mobile communication facility 102 and in part on a server. The predictive text may be based, in part, on at least one of mobile communication facility 102 information, SMS conversion, mobile subscriber characteristic information, **user specific dictionary**, numeric keypad conversion, and multi-letter keypad conversion.

(Ramer, 89:46-52, 92:26-38.)

515. Ramer further discloses that the speech patterns are detected and saved to assist in future interpretation of utterances:

The speech patterns of a unique user may be stored on a hard drive (locally or remotely) or other memory device, and may be loaded into memory, in whole or in part, when the program is run. A comparator may use, for example, correlation or other discrete Fourier transform or statistical techniques to compare the stored patterns against the output of the analog-digital converter.

(Ramer, 21:33-39.) Ramer therefore uses the “*history of a current dialog and one or more prior dialogs*” to interpret a user’s natural language utterance.

516. The combination of Ramer and Cross discloses “*wherein the speech recognition engine recognizes the one or more words or phrases in the natural language utterance using a plurality of entries in one or more dictionary and phrase*

*tables that are dynamically updated based on a history of a current dialog and one or more prior dialogs.”*

**M. Dependent Claims 11, 26, 37 and 52**

11. The method of claim 10, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

26. The method of claim 25, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

37. The system of claim 36, wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

52. The system of claim 51 wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.

517. Dependent claims 11, 26, 37 and 52 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 11, 26, 37 and 52 add the following limitation: “*wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.*”

518. The combination of Ramer and Cross discloses these limitations. Specifically, Cross explains that the dictionary can be created “by selecting words having usage times no earlier than a predetermined threshold ... for example, select only words that have been used within the last five minutes, thereby improving the relevance of words used to supplement (306) the grammar (104).” (Cross, 15:46-53.) By limiting the dictionary to only terms uttered in the last five minutes, Cross (i) determines that words uttered more than five minutes ago are less likely to be relevant to the current utterance and (ii) “derives” the probabilities based on the current dialog since the current dialog sets one bookend to the five minute threshold. The combination of Ramer and Cross discloses “*wherein the plurality of entries in the one or more dictionary and phrase tables are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the prior dialogs.*”

**N. Dependent Claims 12 and 38**

12. The method of claim 1, further comprising determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

38. The system of claim 27, further comprising an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.

519. Claims 12 and 38 depend from claims 1 and 27, respectively, and add the limitation “*further comprising determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.*”

520. The limitation added in claims 12 and 38 is the exact same limitation that is addressed in §VIII.C.7 regarding claim limitation [14F] above. For the same reasons discussed above in §VIII.C.7 for claim limitation [14F], (i) the inventors admitted this limitation was well known in the prior art; and (ii) the combination of Ramer and Cross discloses “*determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive*

*misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event.”*

**O. Dependent Claims 13, 15, 39 and 41**

13. The method of claim 12, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

15. The method of claim 14, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

39. The system of claim 38, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.

41. The system of claim 40, wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the natural language utterance, a subsequent input that overrides the request in a time shorter than

|                                                                                                             |
|-------------------------------------------------------------------------------------------------------------|
| an expected time to process the request, or a subsequent input that repeats the natural language utterance. |
|-------------------------------------------------------------------------------------------------------------|

521. Dependent claims 13, 15, 39 and 41 each add the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Each of claims 10, 25, 36 and 51 add the following limitation: “*wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.*”

522. As discussed above, I understand from the ’176 patent that the inventors were aware that the user of an adaptive misrecognition engine was well known in the art and was not novel. (’176 patent, 4:55-64.) A POSITA would understand that the four types of “*predetermined events*” listed in this claim limitation (“*[i] a subsequent input proximate in time to the natural language utterance, [ii] a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, [iii] a subsequent input that overrides the request in a time shorter than an expected time to process the request, or [iv] a subsequent input that repeats the natural language utterance*”) are typical predetermined events that are regularly detected by adaptive misrecognition

engines. Therefore, a POSITA would understand that the inventors conceded that this limitation was already well known in the prior art.

523. As discussed above related to claim limitation [14F], Ramer discloses “*predetermined events*” that are detected by the “adaptive misrecognition engine” that cause the conversational language processor to reinterpret the words or phrases. Specifically, Ramer discloses a predetermined event of “*receiving one or more of a subsequent input proximate in time to the natural language utterance.*” Ramer discloses that “[a]s the user amends his query entry 120, the cache 2300 may implicitly transfer the content to the mobile search host facilities 114 for real time processing within the disambiguation facility 140, search facility 142, or other facility within the mobile host search facilities 114.” (Ramer, 50:65-51:3.)

524. A POSITA would understand that Ramer disclosing a user “amend[ing] his query” describes a type of predetermined event in which the system “*receiv[es] one or more of a subsequent input proximate in time to the natural language utterance,*” because the system is transferring the user’s amended query “for real time processing within the disambiguation facility,” which a POSITA would understand to mean that subsequent inputs proximate in time to the natural language utterance are used to reinterpret the natural language utterance.

525. Ramer likewise discloses that, when an error is detected, “suggestions related to the query ... may be produced for the user” and once the corrected search

query is determined, “it may then be used to perform a search for results.” (Ramer, 132:42-53.) Here, the predetermined event is the suggestion provided to the user to amend the query. A POSITA would understand that the “corrected search query” is presented “*proximate in time to the natural language utterance.*”

526. The combination of Ramer and Cross discloses “*wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance, a subsequent input that stops the conversational language processor from processing the request contained in the, natural language utterance, a subsequent input that overrides the request in a time shorter than an expected time to process the request, or a subsequent input that repeats the natural language utterance.*”

**P. Dependent Claims 16 and 42**

**1. Preamble [16P] / [42P]**

[16P] The method of claim 14, wherein recognizing the words or phrases in the natural language utterance includes:

[42P] The system of claim 40, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to:

527. Dependent claims 16 and 42 each add nearly the exact same limitation over the independent claims each depends from and I therefore have analyzed them together. Claims 16 and 42 add the following limitations, respectively: “*wherein recognizing the words or phrases in the natural language utterance includes,*” and

*“wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to.”*

528. The preambles for claims 16 and 42 simply refer back to the claims from which they depend and do not add any substantive limitations. As discussed above in §VIII.C regarding claims 14 and 40, those claims are all disclosed in the prior art and would have been well known to a POSITA.

## **2. Mapping a Stream of Phonemes [16A] / [42A]**

[16A] mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

[42A] map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar; and

529. Claims 16 and 42 add the following limitations, respectively: *“mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar,”* and *“map a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar.”*

530. The combination of Ramer and Cross discloses *“mapping a stream of phonemes contained in the natural language utterance to one or more syllables that are phonemically represented in an acoustic grammar,”* and *“map a stream of phonemes contained in the natural language utterance to one or more syllables that*

*are phonemically represented in an acoustic grammar.”* Claim limitations [16A] and [42A] are the same as those for claim limitations [1C] and [27C], respectively. For the same reasons discussed above in §VIII.B.4, the combination of Ramer and Cross discloses these limitations.

### **3. Generating a Preliminary Response [16B] / [42B]**

[16B] generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.

[42B] generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.

531. Dependent claims 16 and 42 add the following limitations, respectively: “*generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases,*” and “*generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*”

The combination of Ramer and Cross discloses “*generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables*

*includes the recognized words or phrases,” and “generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.”* Claim limitations [16B] and [42B] are the same as those for claim limitations [1D] and [27D], respectively. For the same reasons discussed above in §VIII.B.5, the combination of Ramer and Cross discloses these limitations.

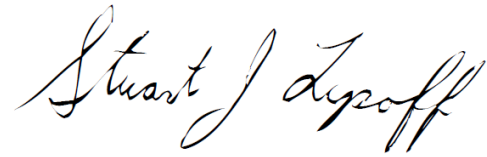
## **IX. Conclusion**

532. In signing this Declaration, I recognize that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in this proceeding. If required, I will appear for cross-examination at the appropriate time. I reserve the right to offer opinions relevant to the invalidity of the challenged claims at issue and/or offer testimony in support of this Declaration.

I hereby declare that all statements made herein of my own knowledge are true and that all statements are 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 18 U.S.C. § 1001.

Dated: April 16, 2025

Respectfully submitted,

A handwritten signature in black ink, reading "Stuart J. Lipoff". The signature is written in a cursive style with a large, stylized 'S' and 'L'.

Stuart Lipoff

## APPENDIX

| Exh. | Reference                                                                                                                                                                                                               |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001 | U.S. Patent 7,818,176 to Freeman et al. (“the ’176 patent”)                                                                                                                                                             |
| 1002 | Prosecution History for U.S. Patent 7,818,176 (“’176 PH”)                                                                                                                                                               |
| 1005 | WO2004/055781 to Taschereau et al. (“Taschereau”)                                                                                                                                                                       |
| 1006 | U.S. Patent 5,640,490 to Hansen et al. (“Hansen”)                                                                                                                                                                       |
| 1007 | U.S. Publication 2010/0125458 to Franco et al. (“Franco”)                                                                                                                                                               |
| 1008 | U.S. Publication 2007/0174258 to Jones et al. (“Jones”)                                                                                                                                                                 |
| 1009 | U.S. Patent 7,577,665 to Ramer et al. (“Ramer”)                                                                                                                                                                         |
| 1010 | U.S. Patent 8,332,218 to Cross et al. (“Cross”)                                                                                                                                                                         |
| 1011 | Memorandum Order Regarding Claim Construction from <i>VB Assets, LLC v. Amazon.com Servs. LLC</i> , No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019) (“Amazon Litigation Mem. Op. Regarding Claim Construction”)     |
| 1012 | Second Amended Joint Claim Construction Chart from <i>VB Assets, LLC v. Amazon.com Servs. LLC</i> , No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019) (“Amazon Litigation Second Am. Joint Claim Construction Chart”) |
| 1013 | Trial Transcript, Volume 1 from <i>VB Assets, LLC v. Amazon.com Servs. LLC</i> , No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019) (“Amazon Litigation Trial Tr. Vol. 1”)                                             |
| 1014 | Petition for <i>Inter Partes</i> Review of U.S. Patent 7,818,176 (Paper 1) from <i>Amazon.com, Inc. v. VB Assets, LLC</i> , IPR2020-01390 (P.T.A.B. July 29, 2020) (“Amazon-VBA-IPR Petition”)                          |
| 1015 | Decision Denying Institution (Paper 7) from <i>Amazon.com, Inc. v. VB Assets, LLC</i> , IPR2020-01390 (P.T.A.B. Mar. 11, 2021) (“Amazon-VBA-IPR Decision”)                                                              |

| Exh. | Reference                                                                                                                                                                                                                                                        |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1016 | U.S. Publication 2007/0067174A1 to Verma et al. (“Verma”)                                                                                                                                                                                                        |
| 1017 | U.S. Patent 6,804,330 to Jones et al. (“Jones ’330”)                                                                                                                                                                                                             |
| 1018 | U.S. Patent 7,620,549 to Di Cristo et al. (“Di Cristo”)                                                                                                                                                                                                          |
| 1019 | Niels Ole Bernsen et al., <i>Designing Interactive Speech Systems: From First Ideas to User Testing</i> (1998) (“Bernsen”)                                                                                                                                       |
| 1020 | Xuedong Huang et al., <i>Spoken Language Processing</i> (2001) (“Huang”)                                                                                                                                                                                         |
| 1021 | Michael F. McTear, <i>Spoken Dialogue Technology</i> (2004) (“McTear”)                                                                                                                                                                                           |
| 1022 | Andrew Kehler, <i>The Effect of Establishing Coherence in Ellipsis and Anaphora Resolution</i> , 31st Annual Meeting of the Association for Computational Linguistics (1993) (“Kehler”)                                                                          |
| 1023 | Sebastian Möller, <i>Quality of Telephone-Based Spoken Dialogue Systems</i> (2005) (“Möller”)                                                                                                                                                                    |
| 1024 | Honghua (Kathy) Dai et al., <i>Detecting Online Commercial Intention (OCI)</i> , Proceedings of the 15th International Conference on World Wide Web (2006) (“Dai”)                                                                                               |
| 1025 | Interactive Advertising Bureau, <i>IAB Internet Advertising Revenue Report</i> (May 2007) (“IAB”)                                                                                                                                                                |
| 1026 | Joe Plummer et al., <i>The Online Advertising Playbook</i> (2007) (“Plummer”)                                                                                                                                                                                    |
| 1027 | Ganesh Iyer et al., <i>The Targeting of Advertising</i> , INSEAD Working Paper Series (Feb. 2004) (“Iyer”)                                                                                                                                                       |
| 1028 | Alex Kinnier, <i>Why We’re Buying DoubleClick</i> , Google Blog (June 26, 2007), <a href="https://googleblog.blogspot.com/2007/06/why-were-buying-doubleclick.html">https://googleblog.blogspot.com/2007/06/why-were-buying-doubleclick.html</a> (“Google Blog”) |
| 1029 | Daniel C. Fain & Jan O. Pedersen, <i>Sponsored Search: A Brief History</i> , 32(2) Bulletin of the American Society for Information Science & Technology 12 (2006) (“Fain”)                                                                                      |

| Exh. | Reference                                                                                                                                                                                                                                                                           |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1030 | Press Release, Google, Google Launches Self-Service Advertising Program (Oct. 23, 2000), <a href="http://googlepress.blogspot.com/2000/10/google-launches-self-service.html">http://googlepress.blogspot.com/2000/10/google-launches-self-service.html</a> (“Google Press Release”) |
| 1031 | Eric J. Sinrod, Duane Morris LLP, <i>Internet Advertising Via “Behavioral Targeting”</i> , USA Today.com (Sept. 28, 2004) (“Duane Morris”)                                                                                                                                          |
| 1032 | Kate Burns, <i>Ten Golden Rules to Search Advertising</i> , 6(3) Interactive Marketing 248 (2005) (“Burns”)                                                                                                                                                                         |