

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

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.	Grounds for Standing.....	1
II.	Identification of Challenge	1
	A. Prior Art.....	1
	B. Grounds For Challenge	2
III.	The '176 Patent.....	2
	A. Patent Overview	2
	B. Technical Background.....	5
	C. Level of Ordinary Skill	6
	D. Claim Construction	6
IV.	GROUND 1: TASCHEREAU-HANSEN COMBINATION RENDERS OBVIOUS CLAIMS 1-2, 9-11, 27-28, 35-37.....	8
	A. Combination Overview	8
	1. Taschereau	8
	2. Hansen.....	9
	3. Motivation to Combine	10
	B. Independent Claims 1 and 27	12
	1. Preambles [1P]/[27P]	12
	2. Receiving Natural Language Request [1A]/[27A]	12
	3. Recognizing One or More Words in the Utterance [1B]/[27B].....	13
	4. Mapping Phonemes [1C]/[27C].....	15
	5. Generating Preliminary Interpretation of Utterance [1D]/[27D]	18

6.	Interpreting the Utterance, Including Establishing Context [1E]/[27E].....	22
7.	Selecting An Advertisement [1F]/[27F]	26
8.	Presenting the Selected Advertisement [1G]/[27G]	27
C.	Claims 2, 28.....	28
D.	Claims 9, 35.....	29
E.	Claims 10, 36.....	30
F.	Claims 11, 37.....	34
V.	GROUND 2: TASCHEREAU-HANSEN-FRANCO COMBINATION RENDERS OBVIOUS CLAIMS 12-18, 25-26, 38-44, 51-52	34
A.	Combination Overview	35
1.	Franco.....	35
2.	Motivation to Combine	36
B.	Independent Claims 14, 40.....	37
1.	[14P]-[14E]/[40P]-[40E].....	37
2.	Determining An Incorrect Interpretation Occurred and Reinterpreting [14F]/[40F].....	37
C.	Claims 12, 38.....	43
D.	Claims 13, 15, 39 and 41	43
E.	Claims 16-18, 25, 26, 42-44, 51, 52	44
VI.	GROUND 3: TASCHEREAU-HANSEN-JONES COMBINATION RENDERS OBVIOUS CLAIMS 3-8, 19-24, 29-34, 45-50	45
A.	Combination Overview	45
1.	Jones.....	45

2.	Motivation to Combine	46
B.	Claims 3, 19, 29, 45.....	47
1.	Tracking an Interaction Pattern [3A]/[19A]/[29A]/[45A]	47
2.	Updating a Model Based on the Interaction Pattern [3B]/[19B]/[29B]/[45B]	48
C.	Claims 4, 20, 30, 46.....	49
D.	Claims 5, 21, 31, 47.....	50
E.	Claims 6, 22, 32, 48.....	51
1.	Building Long-Term and Short-Term Knowledge [6A]/[22A]/[32A]/[48A]	51
2.	Receiving a Subsequent Natural Language Utterance [6B]/[22B]	52
3.	Interpreting Using Long-Term and Short-Term Knowledge [6C]/[22C]/[32B]/[48B]	52
F.	Claims 7, 23, 33, 49.....	53
G.	Claims 8, 24, 34, 50.....	54
VII.	GROUND 4: RAMER-CROSS COMBINATION RENDERS OBVIOUS CLAIMS 1-52	54
A.	Combination Overview	54
1.	Ramer	54
2.	Cross.....	56
3.	Motivation to Combine	56
B.	Independent Claims 1 and 27	58
1.	Preambles [1P]/[27P]	58
2.	Receiving Natural Language Request [1A]/[27A]	58

3.	Recognizing One or More Words in the Utterance [1B]/[27B]	59
4.	Mapping Phonemes [1C]/[27C]	60
5.	Generating Preliminary Interpretation of Utterance [1D]/[27D]	62
6.	Interpreting the Utterance Using Context [1E]/[27E]	62
7.	Selecting An Advertisement [1F]/[27F]	64
8.	Presenting the Selected Advertisement [1G]/[27G]	65
C.	Independent Claims 14, 40	66
1.	[14P]-[14E]/[40P]-[40E]	66
2.	Determining An Incorrect Interpretation Occurred and Reinterpreting [14F]/[40F]	66
D.	Claims 2, 18, 28, 44	68
E.	Claims 3, 19, 29, 45	69
1.	Tracking an Interaction Pattern [3A]/[19A]/[29A]/[45A]	69
2.	Updating a Model Based on the Interaction Pattern [3B]/[19B]/[29B]/[45B]	70
F.	Claims 4, 20, 30, 46	71
G.	Claims 5, 21, 31, 47	72
H.	Claims 6, 22, 32, 48	72
1.	Building Long-Term and Short-Term Knowledge [6A]/[22A]/[32A]/[48A]	72
2.	Receiving a Subsequent Natural Language Utterance [6B]/[22B]	74
3.	Interpreting Using Long-Term and Short-Term Knowledge [6C]/[22C]/[32B]/[48B]	75

I.	Claims 7, 23, 33, 49.....	75
J.	Claims 8, 24, 34, 50.....	76
K.	Claims 9, 17, 35, 43.....	76
L.	Claims 10, 25, 36, 51.....	77
M.	Claims 11, 26, 37, 52.....	78
N.	Claims 12, 38.....	79
O.	Claims 13, 15, 39, 41.....	79
P.	Claims 16 and 42.....	80
VIII.	DISCRETIONARY DENIAL IS NOT APPROPRIATE	80
A.	35 U.S.C. §314(a).....	80
B.	35 U.S.C. §325(d).....	81
IX.	MANDATORY NOTICES	81
A.	Real Party In Interest.....	81
B.	Related Matters.....	81
C.	Notice of Counsel and Service Information.....	82
X.	CONCLUSION.....	83

EXHIBIT LIST

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”)
1003	Declaration of Stuart Lipoff in Support of Inter Partes Review of U.S. Patent 7,818,176 (“Lipoff Declaration”)
1004	Curriculum Vitae of Stuart Lipoff (“Lipoff CV”)
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”)

Exh.	Reference
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”)
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)
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), https://googleblog.blogspot.com/2007/06/why-were-buying-doubleclick.html (“Google Blog”)

Exh.	Reference
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”)
1030	Press Release, Google, Google Launches Self-Service Advertising Program (Oct. 23, 2000), http://googlepress.blogspot.com/2000/10/google-launches-self-service.html (“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”)

Samsung Electronics Co. Ltd. and Samsung Electronics America, Inc.

(“Petitioners”) petition for *inter partes* review of claims 1-52 of U.S. Patent 7,818,176 (“the ’176 patent”; EX-1001).

I. Grounds for Standing

Petitioners certify the ’176 patent is available for IPR; Petitioners are not barred or estopped.

II. Identification of Challenge

A. Prior Art

The ’176 patent was filed on February 6, 2007. Each applied reference was filed or published before February 6, 2007.

1. International Publication WO2004/055781 to Taschereau et al.

(“Taschereau”; EX-1005), published July 1, 2004, is prior art under 35 U.S.C. §102(b).

2. **U.S. Patent 5,640,490 to Hansen et al.** (“Hansen”; EX-1006), issued on June 17, 1997, is prior art under 35 U.S.C. §102(b).

3. **U.S. Publication 2010/0125458 to Franco et al.** (“Franco”; EX-1007), filed July 13, 2006, is prior art under 35 U.S.C. §102(e).

4. **U.S. Publication 2007/0174258 to Jones et al.** (“Jones”; EX-1008), filed December 29, 2006, is prior art under 35 U.S.C. §102(e).

5. **U.S. Patent 7,577,665 to Ramer et. al.** (“Ramer”; EX-1009) filed January 19, 2006, is prior art under 35 U.S.C. §102(e).

6. **U.S. Patent 8,332,218 to Cross et. al.** (“Cross”; EX-1010) filed June 13, 2006, is prior art under 35 U.S.C. §102(e).

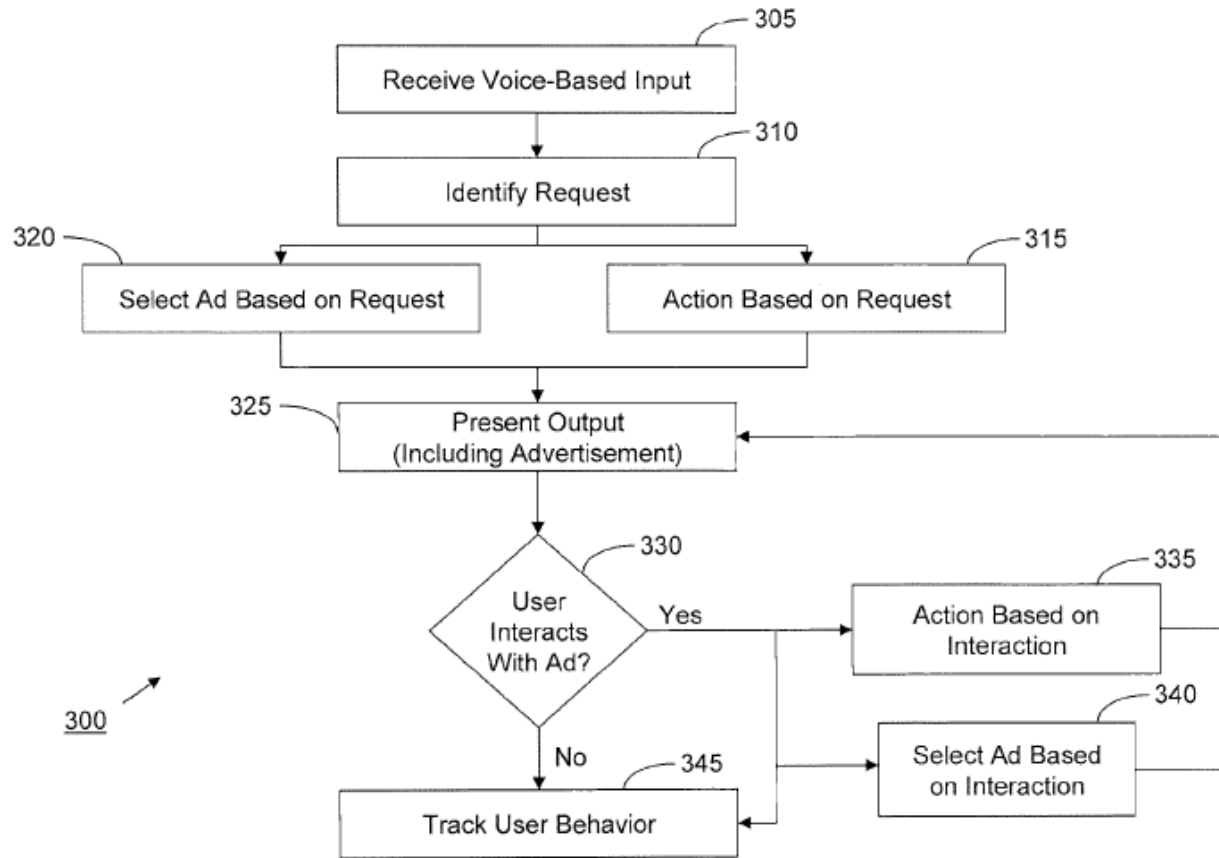
B. Grounds For Challenge

Ground		Claims	Prior Art
1	103	1-2, 9-11, 27-28, 35-37	Taschereau and Hansen
2	103	12-18, 25-26, 38-44, 51-52	Taschereau, Hansen, and Franco
3	103	3-8, 19-24, 29-34, 45-50	Taschereau, Hansen, and Jones
4	103	1-52	Ramer and Cross

III. The '176 Patent

A. Patent Overview

The '176 patent discloses a system, illustrated in Figure 3 (below), for “selecting and presenting advertisements based on natural language processing of voice-based input.” (EX-1001, 1:8-10; 6:66-7:1.)



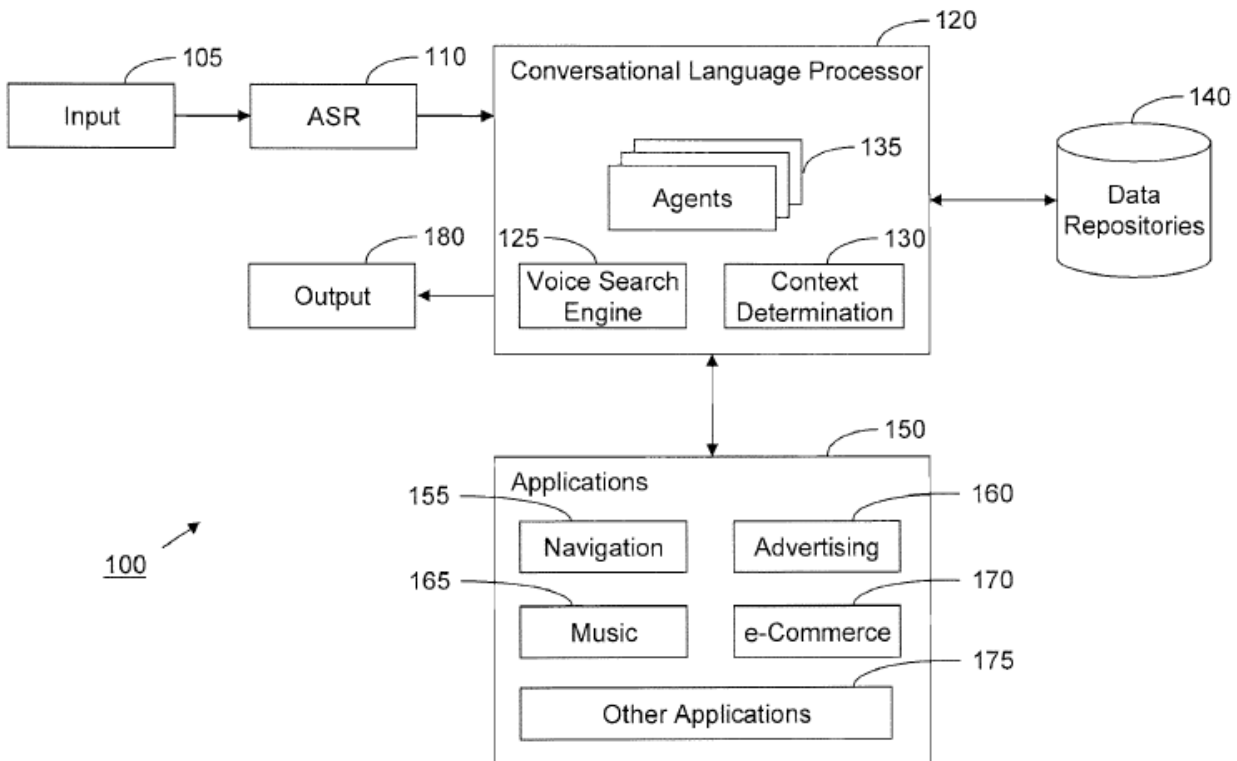
‘176 patent, Figure 3

In “operation 305, [] a voice-based input, including at least a user utterance” is received at a “voice-user interface.” (EX-1001, 7:1-4.) Operation 310 identifies the request in the voice-based input. (EX-1001, 7:10-11.)

The requests can seek information, such as a navigation route, or to perform a task, such as placing a telephone call. (EX-1001, 7:11-31.) Operation 320 (i.e., selecting an ad based on the request) may use profiles to “define affinities between advertisements, topics, users, etc. (e.g., based on click-through or conversion rates,

or other tracking information . . .).” (EX-1001, 10:10-13.) The output presented in operation 325 includes the selected advertisement. (EX-1001, 8:17-18.)

Figure 1 is a representative implementation of the voice user interface. (EX-1001, 2:61-63.)



‘176 patent, Figure 1

An “Automatic Speech Recognizer” (“ASR”) processes the utterance received at input device 105 and generates one or more “preliminary interpretations” using “various techniques,” including “phonetic dictation to recognize a stream of phonemes.” (EX-1001, 3:35-37.) The preliminary interpretation(s) pass to conversational language processor 120, (EX-1001, 3:52-

53), which “establish[es] a context for the current utterance” “(e.g., by scoring possible interpretations and selecting a highest scoring interpretation).” (EX-1001, 4:25-34; EX-1003, ¶¶89-115.)

B. Technical Background

Speech dialogue systems were well known and commercially available in fields such as telephone directory assistance/customer service, dictation, and voice assistance before the priority date. Many of these systems relied one, more or all of the following:

- **Speech Recognizer:** Software and hardware for converting the utterance to text.
- **Semantic Analyzer:** Interprets the text to understand meaning of the utterance.
- **Dialogue Manager:** Interprets the meaning of the utterance in broader context using, for instance, dialog history.
- **Response Generator:** Creates natural language replies.
- **Speech Synthesizer:** Converts the response into speech, and transmits it back to the user.

(EX-1003, ¶44.). Further, these systems expressly incorporated context to facilitate interpretation of meaning through the dialog manager.

Likewise, commercially available systems had already been considering context to select appropriate ads in conjunction with speech recognition tools used to access information (e.g., directory assistance or searching the web). (EX-1003, ¶¶62-73.).

C. Level of Ordinary Skill

A person of ordinary skill in the art (“POSITA”) would have had 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. (EX-1003, ¶¶117-18.)

D. Claim Construction

In *VB Assets, LLC v. Amazon.com Servs. LLC*, No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019) (“VBA-Amazon-Litigation”), the Court construed various terms as set forth below. (EX-1011.)

Claim Term	Construction
“model”	“an approximation, representation, or idealization of selected aspects of the structure, behavior, operation, or other characteristics of a real-world process, concept, or system” (EX-1011, 1)
“acoustic grammar”	“grammar of phonotactic rules of the English language that maps phonemes to syllables” (EX-1011, 2)

“speech recognition engine” / “speech recognition”	“software or hardware that recognizes the words or phrases in the natural language utterance” ¹ (EX-1011, 2)
“context”	No construction necessary (EX-1011, 2 (rejecting Amazon’s proposal that the term should mean “subject domain(s) that applies to a user utterance”))

In *Amazon.com, Inc. et al. v. VB Assets, LLC*, IPR2020-01390 (“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.)

The Board need not expressly construe any term for purposes of this proceeding because Petitioners establish the challenged claims are unpatentable under the constructions entered in the VBA-Amazon-Litigation, the Amazon-VBA-IPR and the claims’ plain meaning, where different. *See Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999).

¹ The VBA-Amazon-Litigation court rejected PO’s position that a “speech recognition engine” is “software or hardware that generates one or more preliminary interpretations of a given utterance.”

IV. GROUND 1: TASCHEREAU-HANSEN COMBINATION RENDERS OBVIOUS CLAIMS 1-2, 9-11, 27-28, 35-37

A. Combination Overview

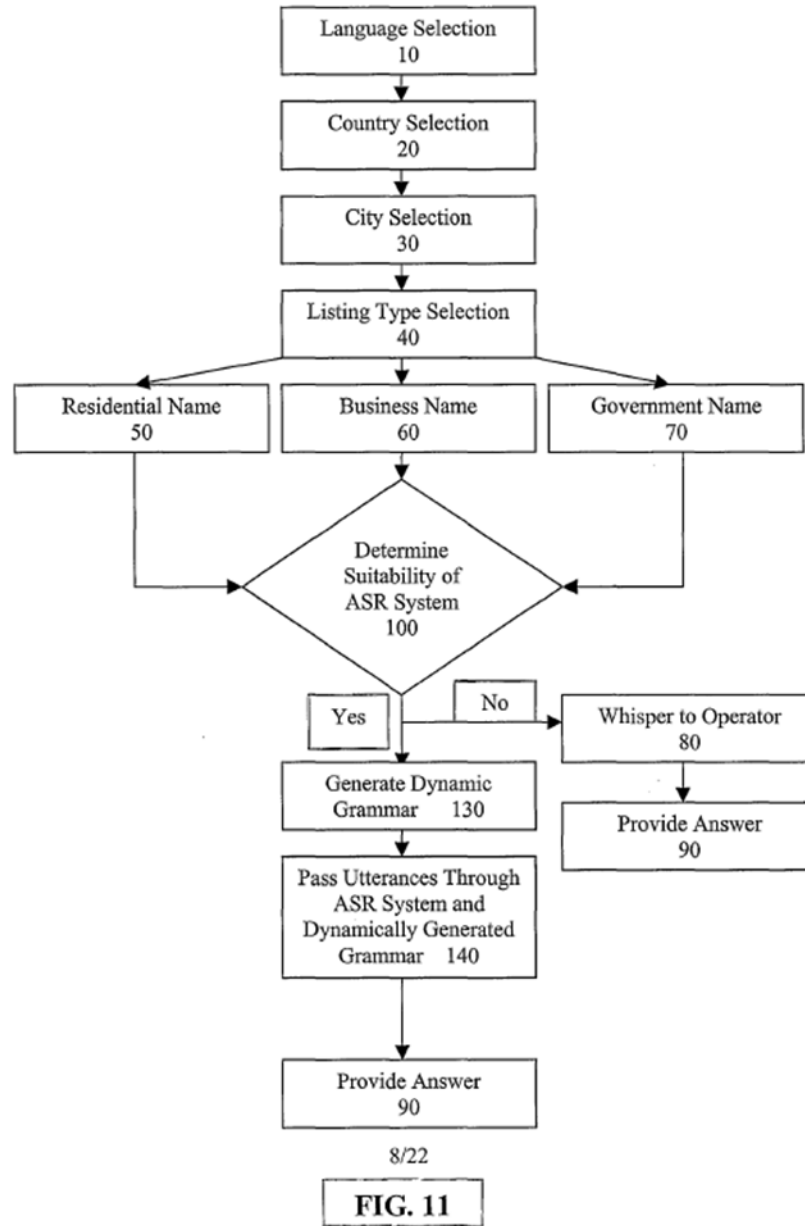
1. Taschereau

Taschereau discloses a speech recognition and response system for telephone directory assistance services. (EX-1005, Abstract.) The system uses an automatic speech recognition (ASR) system which outputs a preliminary interpretation which then passes multiple times through a “dynamically generated grammar” for further contextual analysis. (EX-1005, 7:24-16:3; Fig. 11.)

An example of such contextual analysis is selecting a targeted advertisement to present to directory assistance callers while on hold. “[W]hen 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.”² (EX-1005, 25:5-11; EX-1003, ¶¶137-45.)

Figure 11 depicts an exemplary voice recognition system. (EX-1005, 6:1-2.)

² All emphasis added unless indicated otherwise.



Taschereau, Figure 11

2. Hansen

Hansen discloses “speech recognition” systems (EX-1006, 1:8-10) that “identify[] the phoneme sound types that are contained within an audio speech signal” (EX-1006, Abstract). “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.” (EX-1006, 5:30-35; EX-1003, ¶¶147-52.) A Linguistic Processor generates one or more word possibilities and sends them to a Command Processor which “receives each word, and determines the manner by which it should be used on the host computer.” (EX-1006, 27:38-41; 27:44-49; EX-1003, ¶¶146-59.)

3. Motivation to Combine

A POSITA would have been motivated to incorporate teachings concerning Hansen’s speech recognition implementation into the Taschereau speech recognition system for numerous reasons. First, Taschereau does not disclose how its ASR system converts speech to text. A POSITA would have therefore been motivated to search for references providing such details. A POSITA would have been led to Hansen which discloses well known speech recognition processing techniques. (EX-1003, ¶¶160-61.)

Second, a POSITA would have 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” (EX-1005, 18:6-9; EX-1003, ¶162.)

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) for the above reasons. (EX-1003, ¶163.)

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) into Taschereau’s system of receiving and responding to user’s utterances. Taschereau and Hansen are also 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* EX-1001, 1:66-2:2; EX-1005, 1:6-8, 4:25-27; EX-1006, 27:50-28:3; EX-1003, ¶¶164-65.)

B. Independent Claims 1 and 27

1. Preambles [1P]/[27P]

To the extent the preamble is limiting, the Taschereau-Hansen combination discloses a “*system*” that practices a “*method*” “*for selecting and presenting advertisements in response to processing natural language utterances.*” For example, 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.” (EX-1005, 24:25-26.)

Taschereau describes the system can provide a targeted advertisement based on 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.” (EX-1005, 25:10-11; *see also* claim 9; EX-1003, ¶¶167-70.)

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

The Taschereau-Hansen combination discloses “*receiving a natural language utterance containing at least one request at an input device.*” (EX-1003, ¶¶171-79.) Taschereau discloses that “a user from a phone booth is most likely to be sent to the ASR system.” (EX-1005, 20:24-26.) Here, the microphone in the handset is an “*input device*” that receives the natural language utterance containing a request. (EX-1003, ¶176.) Taschereau discloses that the user’s natural language

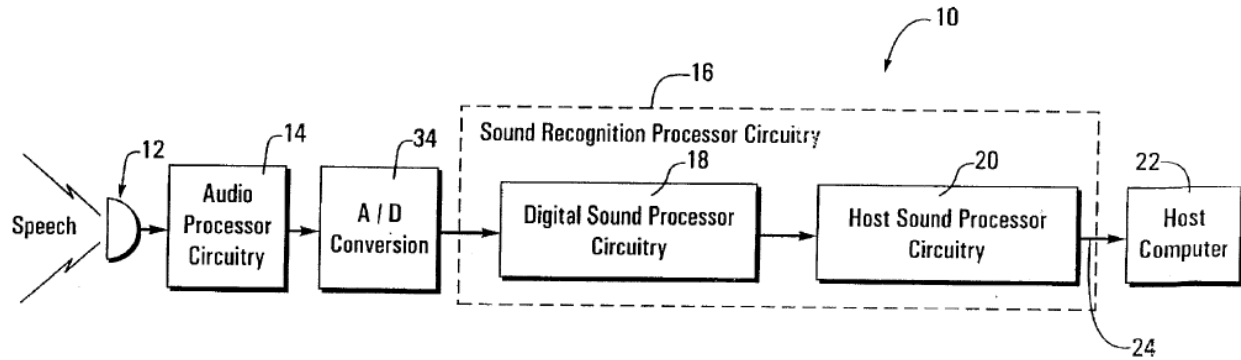
utterance can include a request. (EX-1005, 23:19-21 (“The system and method described above can also serve to direct services to users or direct users to services. For example when a user requests the phone number of a taxi company.”).) (EX-1003, ¶¶173-74.)

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

In Taschereau, “using ASR, speech is simultaneously ‘**recognized**’ and ‘recorded.’” (EX-1005, 13:12-15; *see also* EX-1005, claim 1.) The ASR system receives a natural language utterance and employs four steps (transformation, word map, grammar generation and grammar interpretation) to recognize words or phrases in the utterance. (EX-1005, 7:24-8:1; EX-1003, ¶¶182-83.) Taschereau thus discloses a “*speech recognition engine*” under the VBA-Amazon-Litigation construction.

The phone handset (“*input device*”) is “*coupled*” to a speech recognition engine. (EX-1005, 20:21-26 (discussing use of the ASR system by phone booth users).)

“[S]ystem 10 includes an audio processor means for receiving an audio speech signal and for converting that signal into a representative audio electrical signal.” (EX-1006, 4:36-39; Figure 1 (below).) Audio processor circuitry 14 “receives the audio electrical signal generated by the microphone 12, and then functions so as to condition the signal.” (EX-1006, 4:48-51.)



Hansen, Figure 1

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.” (EX-1006, 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.” (EX-1006, 5:30-35.) Hansen thus discloses a “*speech recognition engine*” under the VBA-Amazon-Litigation construction. (EX-1003, ¶¶185-88.)

The Taschereau-Hansen combination thus 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.*” (EX-1003, ¶¶180-89.)

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

The Taschereau-Hansen combination discloses these limitations.

In Hansen, Sound recognition processor circuitry 16 “identif[ies] the specific phoneme sound types contained within the audio speech signal.” (EX-1006, 4:62-67.)

“Once the sound recognition processor circuitry 16 has extracted the corresponding phoneme sounds, it programmably invokes a series of linguistic program tools” that “translates the series of identified phonemes into the corresponding syllable, word or phrase.” (EX-1006, 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.

(EX-1006, 10:2-11; EX-1003, ¶195.)

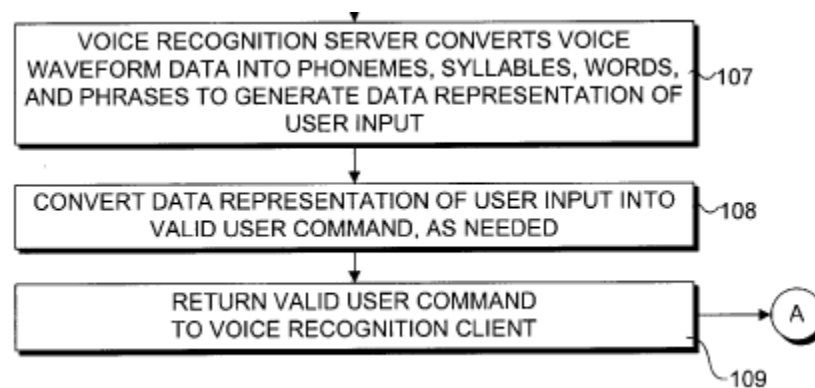
The phonetic dictionary includes English language rules. (EX-1006, 24:39-44 (the phonetic dictionary “contains the **English spelling** of a word along with its corresponding phonetic representation”), *see also* 24:50-52 (the phonetic dictionary “may include a database that contains a more comprehensive Webster-like collection of words”).) The phonetic dictionary maps phonemes to syllables.

(EX-1006, 10:2-11.) Hansen thus discloses an “*acoustic grammar*” because it includes a grammar of phototactic English language rules that maps phonemes to syllables under the VBA-Amazon-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. (EX-1003, ¶196.)

Moreover, phonetic dictionaries mapping phonemes to syllables were well known in the art. For example:

- U.S. Patent Publication No. 2007/0067174 (“Verma”) discloses phonemes of speech utterances mapped to syllables using a syllabic mapping database, which “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” so that “. 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].)
- U.S. Patent 6,804,330 (“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.” (EX-1017, 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.” (EX-1017, 8:27-31.) For example (as shown in Figure 5A, below), 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 “tttt”, which may be chained together to form the syllables “in” and “put.” (EX-1017, 8:35-45.)



Jones '330, Figure 5A (excerpt)

(EX-1003, ¶¶199-200.)

The Taschereau-Hansen combination thus 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.” (EX-1003, ¶¶190-201.)

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

Hansen discloses “*generat[ing] a preliminary interpretation for the natural language utterance from the one or more syllables*” in three stages: (i) Dictionary Lookup, (ii) Word Collocations and (iii) Grammar Check. (EX-1006, Fig. 10; EX-1003, ¶¶202-12).

“[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**.” (EX-1006, 24:39-44.)

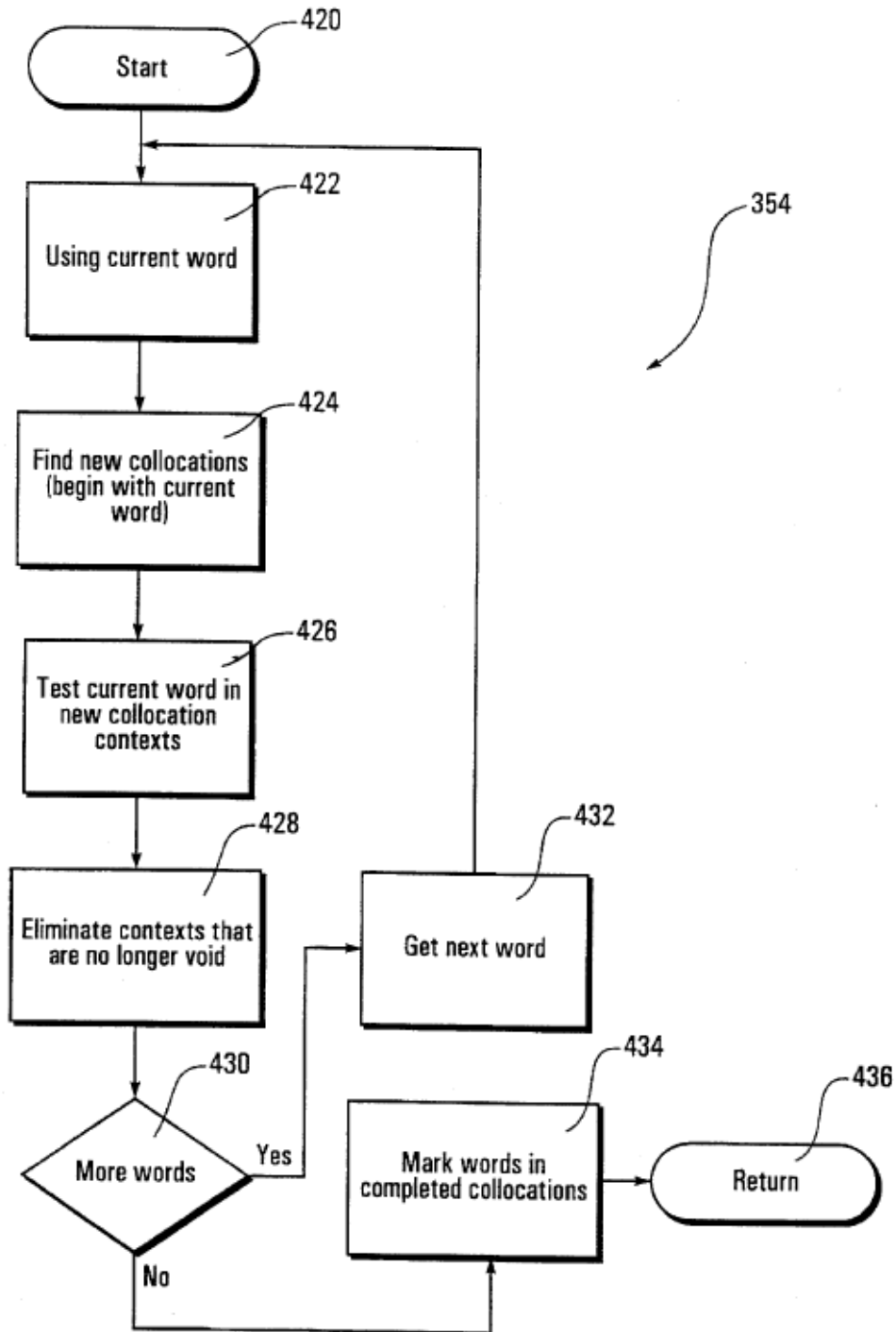
Dictionary Lookup “scans the appropriate dictionary to determine if the incoming sequence of sounds ... form a complete word, or the beginnings of a possible word.” (EX-1006, 24:62-65; Fig. 10A; EX-1003, ¶204.)

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.’” (EX-1006, 25:1-2.) Hansen equates “sequences” with “syllables.” (EX-1006, 25:17-19 (“An ‘active’ sequence is a sequence that could potentially form a word with the addition of a new sound or sounds.”), *see also* 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.”.) Also, 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.” (EX-1017, 8:28-30; EX-1003, ¶205.)

The phonetic dictionary is used to generate a preliminary interpretation of the utterance. (EX-1006, 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 ... for further analysis.”).) This block (354) is called “word collocations.” (EX-1006, Fig. 10; EX-1003, ¶206.)

As shown in Figure 10B, below, “the 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.” (EX-1006, 26:20-23; EX-1003, ¶207.)



Hansen, Figure 10B

“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. ... If such a collocation or collocations exist, then the current word possibility is tested to see if it fits within the collocation context.” (EX-1006, 26:32-42.)

“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.” (EX-1006, 26:53-60.)

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.*” (EX-1003, ¶¶204-10.)

Hansen also discloses that the “*the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases.*” (EX-1003, ¶211.) 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.” (EX-1006, 3:7-12.)

The Taschereau-Hansen combination thus 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.*” (EX-1003, ¶¶202-12.)

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

The Taschereau-Hansen combination discloses these limitations. Taschereau discloses interpreting the recognized words at a “*conversational language processor.*” (EX-1003, ¶¶216-18.) Taschereau discloses an “adaptive” system that “can take advantage of linguistics ... and can extract meaning from the utterance.” (EX-1005, 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.” (EX-1005, 23:7-9.)

Taschereau achieves this goal by interpreting the recognized words and phrases with a multi-pass process using a “dynamically generated ‘latent grammar.’” (EX-1005, 14:7-8.) The “stored utterance” (preliminary interpretation) is “re-input against the latent recognition grammar.” (EX-1005, 15:22-24.) The utterance can be “run through the latent grammar (more than once if necessary) and an answer is provided.” (EX-1005 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.” (EX-1005, 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.” (EX-1005, 21:20-24.) Because the multi-pass latent recognition process recognizes words and phrases, it is a “*conversational language processor*.” (EX-1003, ¶218.)

Taschereau also discloses “*wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance*.” (EX-1003, ¶218.) The ’176 patent gives examples of “context” through reference to U.S. patent application 11/200,164, noting that “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.” (EX-1001, 4:64-5:3.)

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.” (EX-1005, 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.” (EX-1005, 23:24-24:2.)

Thus, a POSITA would have understood that Taschereau’s multi-pass process leveraging a latent-recognition grammar is a “*conversational language processor*” interpreting the recognized words and phrases using grammars broken down into very specific classes such as pizza restaurants and establishing a context for the natural language utterance, such as a taxi request. (EX-1003, ¶220.) Indeed, as discussed below, such location information can be used to select targeted advertisements to present to the user. (*Infra* §IV.B.7.)

Hansen likewise discloses a “*conversational language processor*.” (EX-1003, ¶¶222-25.) “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.’” (EX-1006, 27:34-37.) “[T]he Command Processor is a series of program steps that are executed by a Host Computer 22.” (EX-1006, 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.” (EX-1006, 27:44-49; 27:52-28:11.)

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. (EX-1003, ¶225.)

Hansen also discloses the step “*wherein interpreting the recognized words or phrases includes establishing a context for the natural language utterance.*” (EX-1003, ¶226.) 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.” (EX-1006, 27:44-49.) 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* §IV.A.2.) Each is a “*context for the natural language utterance.*” (EX-1003, ¶226.)

The Taschereau-Hansen combination 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.*” (EX-1003, ¶¶213-27.)

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

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 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.

(EX-1005, 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). (EX-1003, ¶¶228-31.)

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

The Taschereau-Hansen combination discloses “*presenting the selected advertisement via an output device coupled to the conversational language processor.*” (EX-1003, ¶¶232-37.)

As noted in the excerpt above, Taschereau discloses that advertisements are “present[ed]” and 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.” (EX-1005, 25:5-11.)

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.**” (EX-1005, 14:16-19.)

A POSITA would understand 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). (EX-1003, ¶236.)

C. Claims 2, 28

The Taschereau-Hansen combination 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.*” (EX-1003, ¶¶238-45.)

As discussed above, Taschereau discloses a system that uses the interpretation of a user’s natural language request to select and present an advertisement to the user. (*See supra* §§IV.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.” (EX-1005, 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). (EX-1003, ¶¶241-42.)

In addition, a POSITA would have understood that the user location information accessed in 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. (EX-1001, 4:64-5:3.) Taschereau therefore discloses a

“*model*” under the VBA-Amazon-Litigation construction. 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. (EX-1001, 4:64-5:3; EX-1003, ¶¶243-44.)

D. Claims 9, 35

The Taschereau-Hansen combination 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.*” (EX-1003, ¶¶246-49.)

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’.” (EX-1005, 24:6-8.) Put another way, Taschereau describes providing advertisements in response to a user’s request for a business in a particular area. (EX-1005, 25:5-11.) In the example above, Taschereau’s conversational language processor 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.*” (EX-1003, ¶247.)

Taschereau also discloses its system can resolve the user’s request “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.” (EX-1005 24:3-6.) 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. (EX-1003, ¶248.)

E. Claims 10, 36

The Taschereau-Hansen combination discloses “*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.*” (EX-1003, ¶¶250-55.)

As discussed above, the Taschereau-Hansen combination includes a “*speech recognition engine*” that “*recognize[s] one or more words or phrases in a natural language utterance.*” (See *supra* §IV.B.3.)

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. (EX-1005, 7:24-25; 10:21-24.) Grammars are initially generated from the Word List which is a collection of words in a directory that may be identified. (EX-1005, 12:20-21; Abstract.)

A given utterance is then evaluated against the grammar and “n-best” matches are obtained. (EX-1005, 13:26-28.) The “n-best” words are determined based on “[a] list of grammar results and associated return result frequency and confidence scores.” (EX-1005, 13:28-29.) 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

The Word List is scanned and “all entries containing any of the n-best words ... are placed in a dynamically generated ‘latent grammar.’” (EX-1005, 14:8-9.)

“The utterance is then run through the latent grammar (more than once if necessary) and an answer is provided.” (EX-1005, 17:1-12.)

Taschereau 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”³ (which includes the “n-best words”) is the “*plurality of entries in one or more dictionary and phrase tables.*” (EX-1005, 14:26-15:2 (giving exemplary phrase tables).)

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. (EX-1005, 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.” (EX-1005, 13:29-14:1.)

Moreover, Taschereau states that the grammar (which is used to create the list of “n-best words”) “is dynamically generated knowing that the ASR system will be used again to perform one or more latent, and optionally concurrent,

³ A POSITA would understand the “latent grammar” in Taschereau to be a collection of knowledge that can be leveraged in multiple processes including automatic speech recognition (e.g., converting speech to text) and understanding context in the speech. (EX-1003, ¶253.)

recognitions, each latent recognition evaluating the terms from a previous recognition process.” (EX-1005, 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.” (EX-1005, 14:11-12; EX-1003, ¶¶250-55.)

F. Claims 11, 37

The Taschereau-Hansen combination 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.*” (EX-1003, ¶¶256-58.)

As discussed above, a POSITA would have understood that Taschereau’s dynamically updating the “latent grammar”/“n-best words” based on confidence scores are dynamically updating “*based on ... prior probabilities derived from the history of the current dialog and the prior dialogs*” because the system is using prior probabilities (confidence scores) to dynamically update the latent grammars. (EX-1003, ¶257.)

V. GROUND 2: TASCHEREAU-HANSEN-FRANCO COMBINATION RENDERS OBVIOUS CLAIMS 12-18, 25-26, 38-44, 51-52

The Taschereau-Hansen combination discloses all elements of independent claims 14 and 40 except determining that the conversational language processor

incorrectly interpreted words or phrases in response to “an adaptive misrecognition engine detecting a predetermined event wherein the conversational language processor reinterprets the words or phrases.” This concept was known and would have been obvious.

A. Combination Overview

1. Franco

Franco describes methods and systems for “correcting errors in speech recognition systems.” (EX-1007, [0002].) Franco includes multiple embodiments for determining that a word or phrase has been incorrectly interpreted.

One 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. (EX-1007, [0018].) In such scenario, the method combines hypothesized interpretations of the first and second (repeated) utterance to maximize the likelihood of correct interpretation. (EX-1007, [0020] (“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.”).) (EX-1003, ¶¶261-64.)

2. Motivation to Combine

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.

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.” (EX-1005, 2:1-4.) A POSITA would understand Taschereau expressing a motivation to combine its system with systems like Franco, that specialize in correcting errors in speech recognition systems.

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.” (EX-1007, [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.

Third, a POSITA would have been motivated to modify the recursive process in Taschereau (which includes “multiple simultaneous passes of the same

utterance” (EX-1005, 19:20-20:7)) with identified “predetermined events” like a user repeating the same utterance in order to reduce processing time.

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.” (EX-1007, [0027].) Taschereau is implemented through the telephone system, which contains communication channels. (EX-1003, ¶¶265-69.)

B. Independent Claims 14, 40

1. [14P]-[14E]/[40P]-[40E]

The elements in limitations [14P]-[14E] and [40P]-[40E] are substantively identical to those in limitations [1P]-[1B] and [1E]-[1G]/[27P]-[27B] and [27E]-[27G]. For the reasons discussed above in §IV.B, the Taschereau-Hansen-Franco combination discloses these limitations. (EX-1003, ¶¶271-76.)

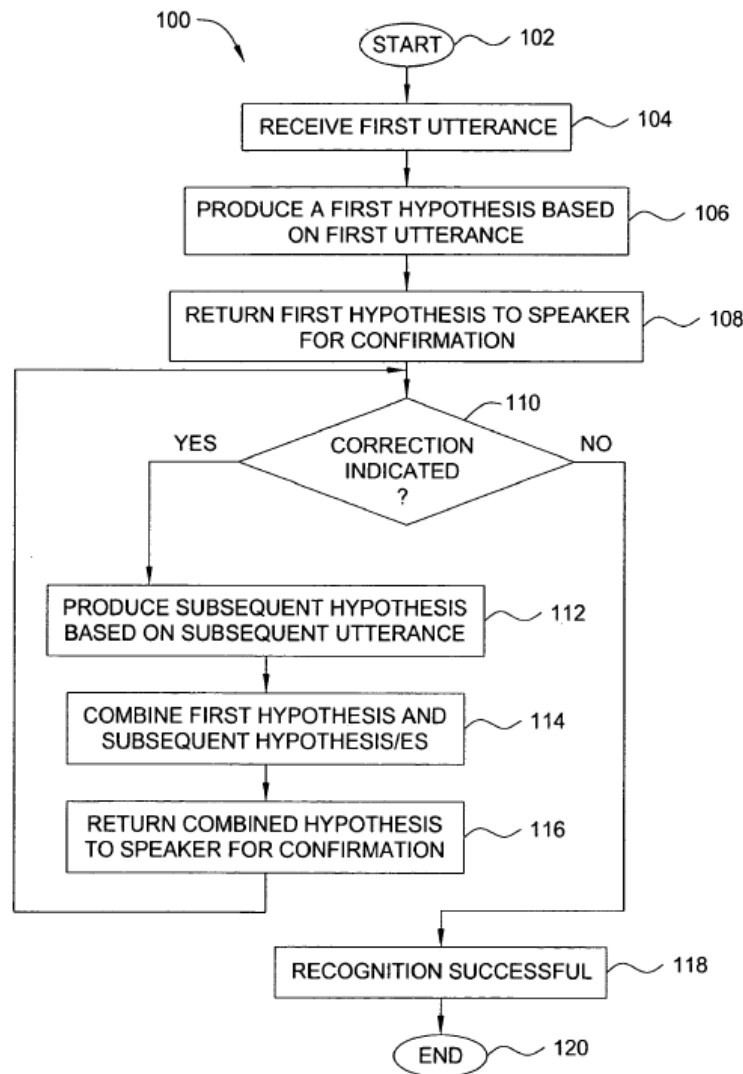
2. Determining An Incorrect Interpretation Occurred and Reinterpreting [14F]/[40F]

The Taschereau-Hansen-Franco combination discloses these limitations.

Franco discloses “*determining that the conversational language processor incorrectly interpret[ing] the words and phrases in response to an adaptive misrecognition engine.*” (EX-1003, ¶¶278-82.) As noted above, in one

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. (EX-1007, [0018], [0020].)

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). (EX-1007, Figure 1.)



Franco, Figure 1

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. (EX-1007, [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*.” (EX-1003, ¶¶281.)

Franco discloses that the incorrect interpretation be made “*in response to an adaptive misrecognition engine detecting a predetermined event*.” (EX-1003, ¶282.)

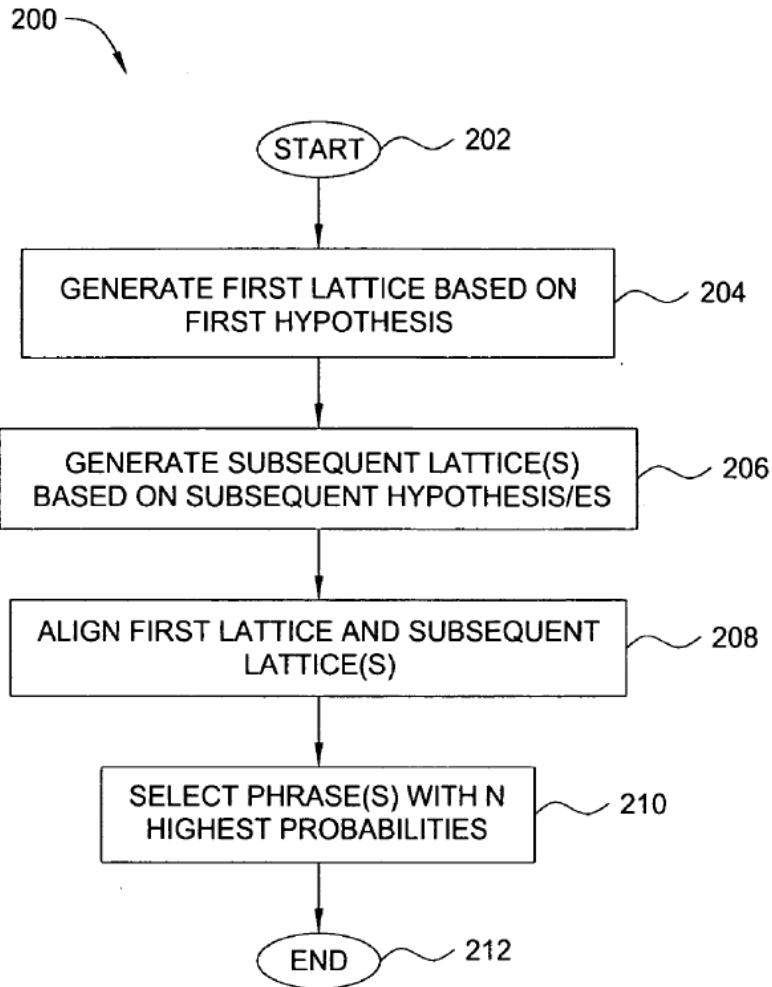
Franco discloses error correction in the specific situation where a speaker naturally repeats an utterance (e.g., “The number is 1234. I repeat, the number is 1234”). (EX-1007, [0018].) Franco’s system assumes that “when speech recognition is unsuccessful on the first attempt, a speaker will naturally repeat all or part of the same utterance in order to correct the recognition error.” (EX-1007, [0012].) Franco likewise discusses error detection when specific keywords are used like “repeating.” (EX-1007, [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. (EX-1003, ¶¶282.)

Franco discloses “*the conversational language processor reinterprets the words or phrases in response to the predetermined event*” in at least two ways. (EX-1003, ¶¶283-89.)

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.” (EX-1007, [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 (EX-1007, [0019]; EX-1003, ¶283.)

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



Franco, Figure 2

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.” (EX-1007, [0026].) Alternatively, the highest probability hypothesis can be presented to the speaker and if the speaker indicates that the hypothesis is incorrect, “then the method 200 selects the phrase with the second highest probability as an alternative combined hypothesis.” (EX-1007, [0026].)

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. (EX-1007, [0020]; EX-1003, ¶¶280-86.)

This disclosure is consistent with the '176 patent intrinsic record. Rather than disclose structure for the “*misrecognition engine*” 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”]. (EX-1001, 4:55-64.) 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. (EX-1018, 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. (EX-1003, ¶¶287-89.)

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.” (EX-1001, 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.” (EX-1001, 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. (EX-1003, ¶290.)

C. Claims 12, 38

Claims 12 and 38 include the same limitation as in [14F] above. For the reasons discussed in §V.B.2, the Taschereau-Hansen-Franco combination discloses these limitations. (EX-1003, ¶¶294-95.)

D. Claims 13, 15, 39 and 41

Franco discloses a predetermined event of “*a subsequent input that repeats the natural language utterance.*” Franco’s system 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.**” (EX-1007, [0012]; *see also* EX-1007, [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.

Franco further discloses a predetermined event of “*a subsequent input proximate in time to the natural language utterance.*” As discussed above, Franco’s method leverages lattice positions or slots. (*Supra* §V.B.2.) “[E]ach slot corresponds to a **temporal interval** in the first utterance. Each slot is associated with one or more words that may potentially have been uttered during the corresponding **temporal interval**. ... 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.” (EX-1007, [0022].)

The Taschereau-Hansen-Franco combination thus discloses “*wherein the predetermined event includes receiving one or more of a subsequent input proximate in time to the natural language utterance ... [and] a subsequent input that repeats the natural language utterance.*” (EX-1003, ¶¶296-301.)

E. Claims 16-18, 25, 26, 42-44, 51, 52

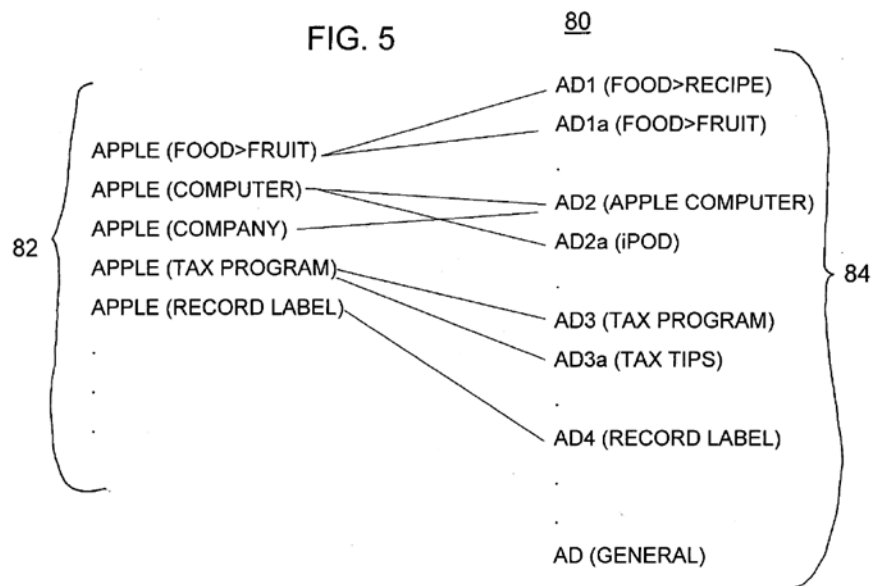
Claims 16 and 42 are substantively identical to limitations [1C]-[1D]/ [27C]-[27D] (§IV.B.4-5). Claims 17 and 43 include the same limitations as in claims 9 and 35 (§IV.D). Claims 18 and 44 include the same limitations as in claims 2 and 28 (§IV.C). Claims 25 and 51 include the same limitations as in claims 10 and 36 (§IV.E). Claims 26 and 52 include the same limitations as in claims 11 and 37 (§IV.F). For the same reasons in the referenced sections, the Taschereau-Hansen-Franco combination discloses these limitations.

VI. GROUND 3: TASCHEREAU-HANSEN-JONES COMBINATION RENDERS OBVIOUS CLAIMS 3-8, 19-24, 29-34, 45-50

A. Combination Overview

1. Jones

In Jones, “[t]argeted advertisement is provided to mobile device users based on one or more keywords spoken by the mobile device users.” (EX-1008, Abstract.) After users “submit voice requests” to a phone-based directory assistance provider, “advertisements and/or products associated with the keyword(s) are provided to the users while the users are waiting for responses to the requests.” (EX-1008, Abstract.) Jones discloses identifying a context associated with an utterance as depicted in Figure 5.



Jones Figure 5

Jones further discloses the use of analytics to track user engagement with advertisements and to select advertisements to be presented. (EX-1008, [0033], [0063]; EX-1003, ¶¶306-11.)

2. Motivation to Combine

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

First, Taschereau expressly provides financial motivation for the combination, noting 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.” (EX-1005, 25:12-15.) A POSITA would have been motivated to incorporate the analytical features disclosed in Jones in order to optimize targeted advertisements to users and maximize revenue. Jones similarly discloses the goal of maximizing advertisement revenue and provides a variety of ways to monetize the advertising feature. (EX-1008, [0062].) For example, Jones discloses tracking user interaction so that income from advertisements can be based on that user interaction. (EX-1003, ¶¶312-14.)

Second, Taschereau and Jones each discuss the importance of being as specific as possible in targeting advertising to users of the telephone directory system. (EX-1005, 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.”); EX-1008, [0005] (“current advertising technologies are directed to mass advertisement and do not provide customized advertisements directed to interests particular to users or their needs”).) A POSITA would understand both Taschereau and Jones to recognize that the more specific and accurate the advertisements are for a particular user, the better the user experience will be. (EX-1003, ¶¶315-16.)

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 each of Taschereau, Hansen and Jones use the same system architecture, namely, the use of speech recognition in conjunction with direct assistance systems. (EX-1003, ¶¶312-17.)

B. Claims 3, 19, 29, 45

1. Tracking an Interaction Pattern [3A]/[19A]/[29A]/[45A]

The Taschereau-Hansen-Jones combination discloses this limitation. 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.

(EX-1008, [0033].) A POSITA would understand each of the foregoing activities to be “*tracking an interaction pattern with the advertisement presented via the output device.*” Jones likewise discloses updating a database to include such tracking information. (EX-1008, [0028].) Jones further discloses collecting and storing geographic information about a user. (EX-1008, [0031]; EX-1003, ¶¶319-22.)

2. Updating a Model Based on the Interaction Pattern [3B]/[19B]/[29B]/[45B]

The Taschereau-Hansen-Jones combination discloses this limitation.

After Jones’s system collects the “interaction pattern tracked for the advertisement,” it stores the information in a database that includes the models which the system uses to select an advertisement. (EX-1008, [0031] (“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 describes an “*environmental model based on the interaction pattern tracked for the advertisement*” because the system selects advertisements that best fit the query and user’s telephone number, which indicates where the user may be located. Jones further discloses that an advertisement may be presented based on “popularity with users,” among other factors. (EX-1008, [0029].)

By updating the Jones database with “user location, user activity, [and] user actions,” 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 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 other users. (EX-1008, [0031]; EX-1003, ¶¶323-28.)

C. Claims 4, 20, 30, 46

The Taschereau-Hansen-Jones combination discloses updating an environmental model by building “*statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.*” As discussed above (§VI.B.2), Jones updates the database to include certain types of information about user behavior or location, which is the “*environmental model.*” In Jones, the environmental model includes statistical profiles for selecting subsequent user advertisements. For example, Jones describes selecting advertisements by referencing a keyword histogram, which is a statistical profile based on prior user interactions and location. (EX-1008, [0031].) The popularity of an advertisement and geographical location are used in determining what advertisement to select and provide to a user. (EX-1008, [0029].) A POSITA would understand that for Jones to know an advertisement is “popular with users”

the system would necessarily have to “*build a statistical profile*” based on prior user interactions. (EX-1003, ¶¶331-33.)

Taschereau likewise 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.” (EX-1005, 25:8-9.) With such information, “very precise advertisements can be presented to the user.” (EX-1005, 25:10-11.) A POSITA would understand that the system would have an “*environmental model to build statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.*” (EX-1003, ¶334.)

D. Claims 5, 21, 31, 47

The Taschereau-Hansen-Jones combination discloses this limitation. The keyword histogram in Jones is used to “*identify affinities between the advertisement presented via the output device and one or more of the recognized words or phrases.*” (EX-1008 [0031] (“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 discloses various ways to determine affinity including “a number of times a particular advertisement is presented in association with one or more keywords.” (EX-1008, [0033]; EX-1003, ¶¶335-38.)

E. Claims 6, 22, 32, 48

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

In Jones, “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.” (EX-1008, [0032].) The database also maintains and uses information regarding user engagement when selecting future advertisements. (EX-1008, [0033].) The system may also select advertisements based on user phone number and/or a user profile, which is “*long-term shared knowledge*” about the user. (EX-1008, [0063]; EX-1003, ¶¶341-42.)

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. (EX-1008, [0031] (“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.”).) (EX-1003, ¶343.)

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. (EX-1008, [0026] (“The system

10 may keep track of the port from which a call originated as well as the telephone number of the caller”).) Such “geographical location of the user” is then used to select an advertisement to be presented. (EX-1008, [0029]; EX-1003, ¶¶344-46.)

2. Receiving a Subsequent Natural Language Utterance [6B]/[22B]

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).” (EX-1008, [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 EX-1008, [0053]; EX-1003, ¶¶347-48.)

3. Interpreting Using Long-Term and Short-Term Knowledge [6C]/[22C]/[32B]/[48B]

As noted above (§VI.E.2), 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 §VI.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*.” (EX-1008, [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*.” (EX-1003, ¶¶349-50.)

F. Claims 7, 23, 33, 49

As noted above (§VI.B.1), Jones discloses “*tracking an interaction pattern with the advertisement*.” 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. (EX-1008, [0028], [0033].) When the user views, clicks on or makes a purchase that interaction pattern is tracked for the advertisement.

Jones also teaches that, after providing an advertisement, the system can offer a coupon related to the advertisement. If the user subsequently requests the coupon, the system will provide the user the coupon. (EX-1008, [0055]-[0056].) A POSITA would have understood 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.*” (EX-1003, ¶¶351-54.)

G. Claims 8, 24, 34, 50

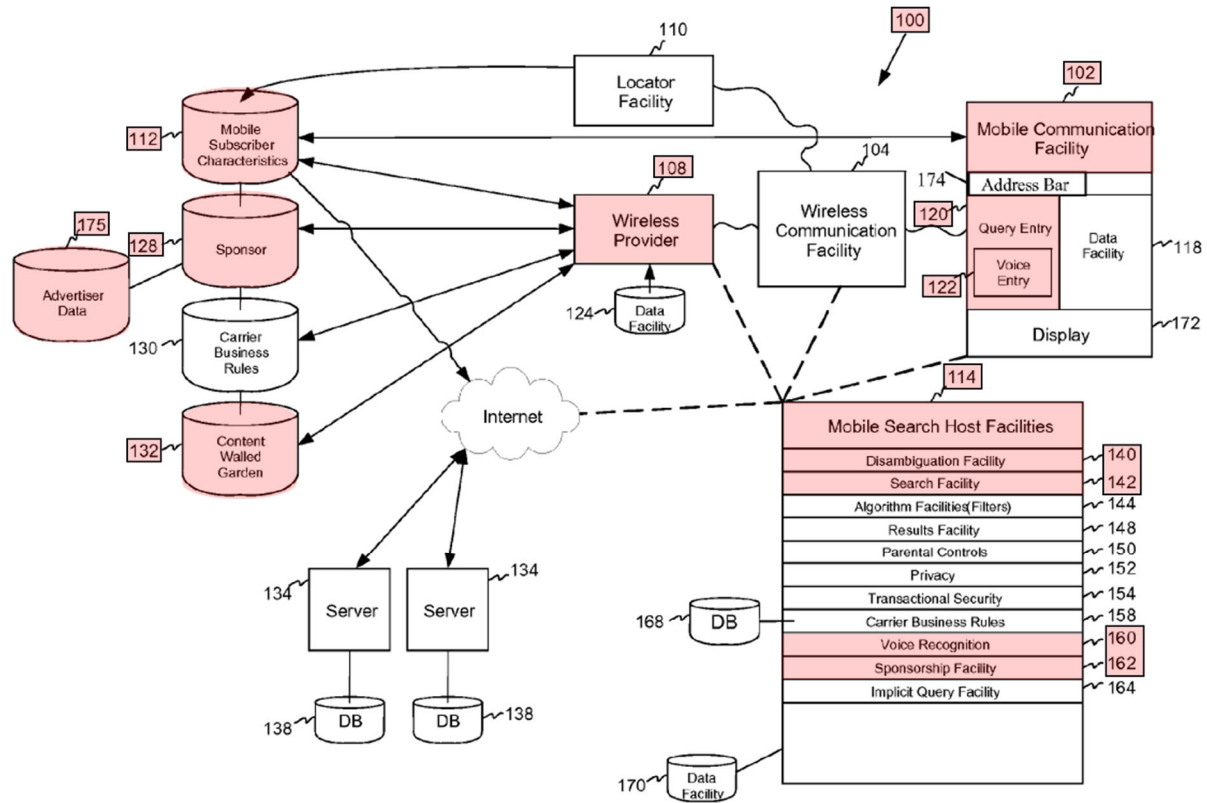
As discussed above (§VI.F), Jones teaches that the system “*execut[es] a task,*” by providing a coupon “*based on the subsequent request that identifies the advertisement,*” the subsequent request being the user’s request for a coupon that identifies the advertisement. (EX-1008, [0055]-[0056]; EX-1003, ¶¶355-56.)

VII. GROUND 4: RAMER-CROSS COMBINATION RENDERS OBVIOUS CLAIMS 1-52

A. Combination Overview

1. Ramer

Ramer discloses (as shown in Figure 1) internet searching with queries input using speech recognition as well as selecting and presenting targeted advertisements to a user conducting a search. (EX-1009, Abstract; 5:6-7.)



Ramer, Figure 1, annotated

The “mobile communication facility” (e.g., mobile phone) includes “speech recognition” functions for inputting information. (EX-1003, ¶¶361-62.) The wireless search platform may include a “data facility containing mobile subscriber characteristics 112 pertaining to individual users,” including “shopping habits,” “click stream information,” “content viewing history” and “transaction history.” (EX-1009, 8:21-48.) Ramer likewise discloses an advertising database for presenting sponsored content to users. (EX-1009, 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.”)) (EX-1003, ¶¶363-70.)

Ramer does not disclose how the speech recognition system is implemented.

A POSITA would have looked to known ways of implementing such a speech recognition system, including those set forth in Cross. (EX-1003, ¶371.)

2. Cross

Cross discloses a “context-based grammar for automatic speech recognition.” (EX-1010, Abstract.) Its “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’), [and] uses the SFV to infer phonemes for the word from a language-specific acoustic model.” (EX-1010, 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.” (EX-1010, 6:38-43.) The system then conducts a contextual analysis to confirm that the word generated by the ASR engine is the right one. (*See, e.g.*, EX-1010, 12:56-64; EX-1003, ¶¶372-78.)

3. Motivation to Combine

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. A

POSITA would have been motivated to make this combination for numerous reasons. First, Ramer suggests the combination when it refers to the user of known voice recognition techniques. (EX-1009, 21:22-45 (“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.”). Second, Cross acknowledges that its speech recognition system is well known. (*See* EX-1010, 9:44-49) (discussing commercially available speech recognition systems which can be modified to add the contextual analysis set forth in Cross).) (EX-1003, ¶¶379-82.)

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. (EX-1003, ¶¶383-85.)

B. Independent Claims 1 and 27

1. Preambles [1P]/[27P]

The Ramer-Cross combination discloses a “*system*” that practices a “*method*” “*for selecting and presenting advertisements in response to processing natural language utterances.*” For example, Ramer 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.” (EX-1009, 83:59-61.) It then processes the natural language utterances and teaches that “recommended advertisements may be paired with webpages presented prior to, during, or following, the display of the advertisement results.” (EX-1009, 2:33-36.)

Ramer thus discloses that the user’s natural language utterance is used for “*selecting and presenting advertisements.*” (EX-1009, 97:52-56.) Such advertisements are selected and presented in “*response to processing natural language utterances,*” (see EX-1009, 83:59-61), and can be based on prior user tracking, (EX-1009, 72:28-37). (EX-1003, ¶¶387-93.)

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

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

Ramer discloses a system that receives natural language utterances containing at least one request. Ramer discloses that “[a] content request may be made by an explicit query request entered by a mobile communication facility user

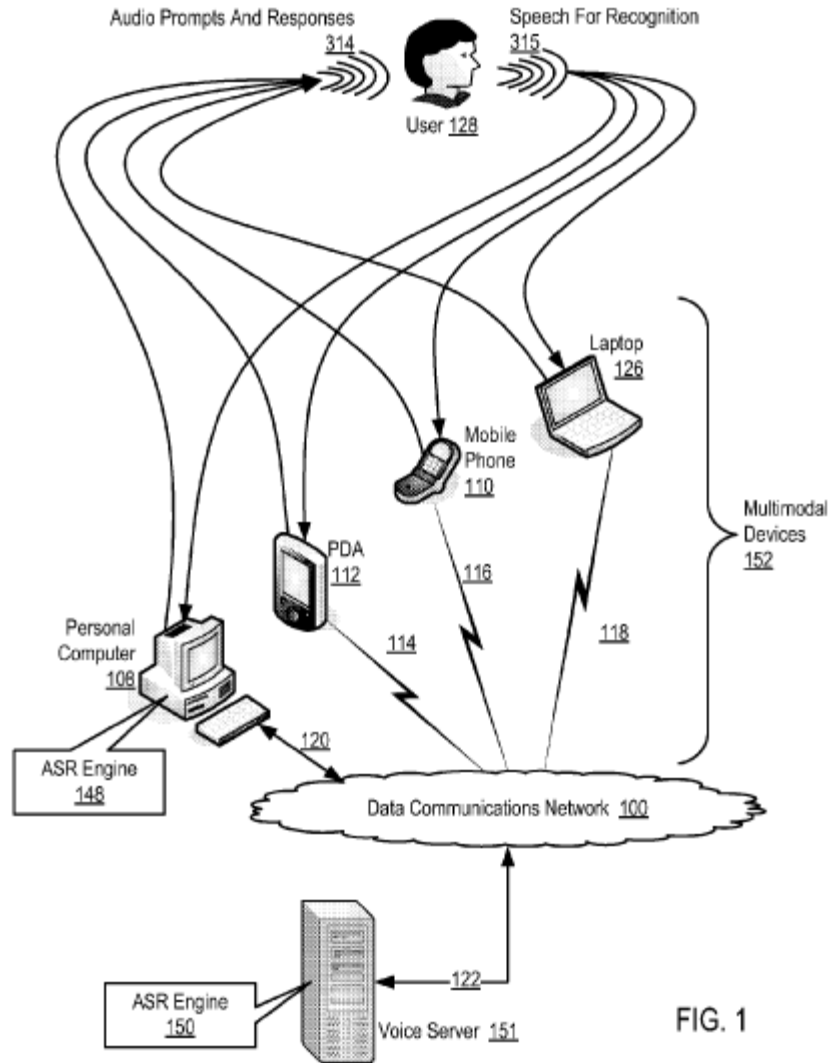
as a voice command in the voice recognition query facility of a mobile communication facility.” (EX-1009, 2:23-26.) Exemplary “mobile communication facilit[ies]” include mobile phones. (EX-1009, 5:22-27.) The mobile phone is an example of an “*input device*.”

Cross likewise discloses this limitation. An exemplary “*natural language utterances containing at least one request*” is: “Show me a quote for an RL.” (EX-1010, 16:50-52.) Like Ramer, such utterances can be received by mobile telephones, (EX-1010, 3:34-47), a type of “*input device*.” (EX-1003, ¶¶394-98.)

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

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

A “multimodal server application (188)” “receive[s] speech for recognition from a user and pass[es] the speech along to an ASR engine (150).” (EX-1010, 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.” (EX-1010, 6:44-46.) Cross thus discloses a “*speech recognition engine*” under the VBA-Amazon-Litigation construction. As shown in Figure 1, the ASR engine is “*coupled to the input device*” (e.g., mobile phone 110). (EX-1003, ¶¶399-404.)



Cross, Figure 1

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

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.”

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.” (EX-1010, 6:33-38.) “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.” (EX-1010, 6:38-43; EX-1003, ¶¶409-11.)

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. (EX-1003, ¶412.) As noted above (*supra* §IV.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,” (EX-1010, 12:1-6), such as English.

Cross thus discloses an “*acoustic grammar*” because it includes a grammar of phototactic English language rules that maps phonemes to syllables under the VBA-Amazon-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. (EX-1003, ¶¶412-14.)

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

Cross discloses *“generat[ing] 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.”*

“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.” (EX-1010, 6:28-33.) As noted above, a POSITA would understand the reference to “pronunciations” to include syllables.

The “ASR engine [] uses the phonemes to find the word in the lexicon.” (EX-1010, 6:42-43.) 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 pronunciations including syllables. (EX-1003, ¶¶415-19.)

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

The Ramer-Cross combination discloses these limitations.

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. (EX-1010, 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”); *see also* Figures 6-10; EX-1003, ¶423.)

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

- “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.” (EX-1010, 17:45-53.)
- “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.”

(EX-1010, 17:53-57.)

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 user information (e.g., location), as context to interpret the recognized words of the natural language utterance. (EX-1009, 24:5-46.)

The Ramer-Cross combination 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.*” (EX-1003, ¶¶420-28.)

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

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. (EX-1009, 46:18-20 (“the bookstore ... may want to advertise to the user 1104 because he is in the area and has some history of searching amazon.com”).)

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 ... 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.” (EX-1009, 47:34-45; 15:53-60.)

The Ramer-Cross combination thus discloses “*selecting an advertisement in the context established for the natural language utterance.*” (EX-1003, ¶¶429-35.)

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

In Ramer, “[t]he 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.” (EX-1009, 48:22-25.) Ramer discloses that the system “[p]resents the selected advertisement via an output device:”

[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.

(EX-1009, 47:34-45.) Because the system receives the ad from the sponsor database, it must be “*coupled to the conversational language processor.*” (EX-1003, ¶¶436-39.)

C. Independent Claims 14, 40

1. [14P]-[14E]/[40P]-[40E]

Limitations [14P]-[14E] and [40P]-[40E] are substantively identical to limitations [1P]-[1B] and [1E]-[1G]/[27P]-[27B] and [27E]-[27G]. For the reasons discussed above in §VII.B, the Ramer-Cross combination discloses these limitations. (EX-1003, ¶¶441-46.)

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

The Ramer-Cross combination discloses these limitations.

Specifically, as shown in Figure 4 below, Ramer includes “disambiguation facility 140” which “may complete or provide more meaning to ambiguous active user inputs,” (EX-1009, 9:14-15), as well as a correction process 244 which creates queries “sufficient to yield intelligible and useful result set(s),” (EX-1009, 24:49-53).

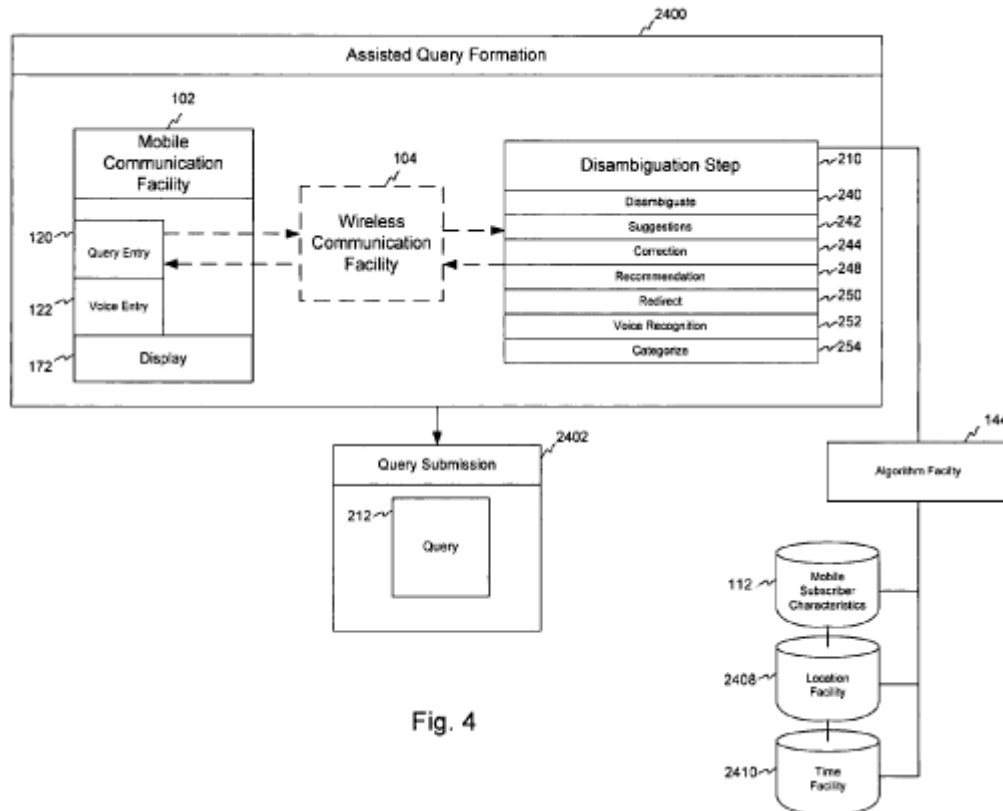


Fig. 4

Ramer, Figure 4

“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.” (EX-1009, 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. (EX-1003, ¶¶450-51.)

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.

In an illustrative example, a user utters “Coltrain” – “not a known word.” (EX-1009, 25:9-15.) The system may misrecognize the word as “Coal Train” or “Soul Train.” (EX-1009, 25:10-11.) The “query assistance facility 210 may predict a correction 244, taking into account user-specific data contained in the purchase history.” (EX-1009, 25:11-13.) Specifically, if there is “a history of many recent purchases of compact discs recorded by the saxophonist, John Coltrane,” the system reinterprets “Coltrain” to “Coltrane.” (EX-1009, 25:3-19.) In doing so, the system “*reinterprets the words or phrases in response to the predetermined event.*” (EX-1003, ¶¶452-53.)

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.” (EX-1009, 25:20-25.)

D. Claims 2, 18, 28, 44

The Ramer-Cross combination discloses these limitations.

Specifically, Ramer discloses a “*conversational language processor [that] elects 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.” First, Ramer discloses selecting an advertisement for a user based on information received from a user’s query. (EX-1009, 48:52-49:4.) Ramer provides an example where a user searches for “Sonny Rollins” (the “*recognized words or phrases*”) and is presented “advertisements and websites for CD’s of Sonny Rollin’s music, videos of his performances, etc.” (EX-1009, 48:52-49:4.)

Second, Ramer discloses selecting an advertisement based on an environmental model. Ramer discloses that “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” and “current time may be used independently or in conjunction with other information to create user recommendations.” (EX-1009, 51:53-59.) As noted above, the user’s location and the time of day are both part of an “*an environmental model derived from environmental conditions or surroundings associated with the specific user.*” (EX-1003, ¶¶456-63.)

E. Claims 3, 19, 29, 45

1. Tracking an Interaction Pattern [3A]/[19A]/[29A]/[45A]

The Ramer-Cross combination discloses this limitation. Ramer teaches that the system tracks “the popularity of a product, service, business, transaction, or website by recording the total number of users in the set that rate it favorably” because the “algorithm assumes that what previous users liked, new users will

like.” (EX-1009, 11:1-11.) Ramer thus discloses “*tracking an interaction pattern with an advertisement.*”

Ramer further discloses a method of “optimizing search results for mobile users may include tracking the on-line interactions” such as “tracking clicks, clickthroughs, queries, [and] clicks following queries.” (EX-1009, 124:5-21; EX-1003, ¶¶465-68.)

2. Updating a Model Based on the Interaction Pattern [3B]/[19B]/[29B]/[45B]

The Ramer-Cross combination discloses this limitation.

After Ramer collects the “*interaction pattern tracked for the advertisement,*” it stores the information in a database along with the models the system uses to select an advertisement.” (EX-1009, 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. (EX-1009, 50:25-35; EX-1003, ¶471.)

Ramer teaches that its algorithm is designed to assume “that what previous users liked, new users will like.” (EX-1009, 11:1-11.) To do so, it must necessarily update its model. Ramer also discloses that “[i]nformation may be stored pertaining to the advertising and web interactions in a database” (EX-1009, 142:10-29) and that the “statistical model may be continually updated to accommodate changing user preferences,” (EX-1009, 139:26-28).

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. For example, the system in Ramer may associate keywords with advertisements that are tied to a specific geographic area or times of day. (EX-1009, 51:49-52:11; EX-1003, ¶¶469-73.)

F. Claims 4, 20, 30, 46

The Ramer-Cross combination discloses updating an environmental model by building “*statistical profiles for selecting subsequent advertisements in response to subsequent natural language utterances.*” As discussed above (§VII.E.2), Ramer stores in the database information about user location and time of day, which is included in the “*environmental model.*”

Ramer discloses that the system uses statistical profiles for selecting subsequent advertisements via an “algorithm, [which] 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.” (EX-1009, 11:12-33; EX-1003, ¶¶476-78.) 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.” (EX-1009, 10:14-18.) Ramer

teaches that “[v]arious weighted average, fuzzy logic, or other techniques may be used to summarize or model” the data. (EX-1009, 10:18-21.)

G. Claims 5, 21, 31, 47

The Ramer-Cross combination discloses this limitation. As discussed above (§VII.E), 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.*” (EX-1009, 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.*” (EX-1009, 10:14-23.) Ramer’s advertisements are designed to “target only a subset of users, such as, only users residing in a particular geographic region, meeting a set of demographic user characteristics.” (EX-1009, 61:23-27; EX-1003, ¶¶479-87.)

H. Claims 6, 22, 32, 48

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

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. For example, Ramer discloses updating the personalized cognitive model as it learns the speech patterns of the user. (EX-1009, 21:22-39.) A POSITA would have understood 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. (EX-1009, 24:63-25:19 (leveraging purchase history and request in current session to identify and fix errors).) (EX-1003, ¶¶489-92.)

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. Ramer discloses building long term information regarding users in response to updating the generalized cognitive model as discussed above in claim 5. Namely, Ramer updates the generalized cognitive model to include information regarding specific information about the user such as age, gender or religion/ cultural affinity. (EX-1009, 26:45-27:27.)

Ramer also discloses a system that can “*build long-term shared knowledge and short-term shared knowledge*” in response to updating an environmental model. As discussed above, Ramer discloses building long term information about

user's geographic location and time (environmental information). (*See supra* §VII.A.1; EX-1009, 25:35-62 (“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.”).)

2. Receiving a Subsequent Natural Language Utterance [6B]/[22B]

As noted above (§VII.H.1), Ramer discloses building long term knowledge about a user's speech pattern to enhance the system's interpretation of user requests. Ramer likewise discloses soliciting and “*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.

(EX-1009, 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). (EX-1003, ¶¶495-96.)

3. Interpreting Using Long-Term and Short-Term Knowledge [6C]/[22C]/[32B]/[48B]

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 one example, as discussed above, Ramer discloses soliciting a subsequent natural language utterance. 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”). (EX-1009, 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. (EX-1009, 24:63-25:19; EX-1003, ¶¶497-98.)

I. Claims 7, 23, 33, 49

As noted above (§VII.E.1), Ramer discloses “*tracking an interaction pattern with the advertisement.*” Ramer also 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.*” (EX-1009, 31:23-42 (“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”).) Ramer provides additional information regarding how the user’s interaction pattern is tracked for advertisements. (EX-1009, 10:31-67.)

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. (EX-1009, 34:4-21; EX-1003, ¶¶499-502.)

J. Claims 8, 24, 34, 50

Ramer’s 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 “sponsored link” after the user specifically identifies the advertisement. (EX-1009, 47:63-48:25) Ramer discloses that clicks and click-throughs are tracked by the system. (EX-1009, 124:13-15.) Therefore, the system disclosed in Ramer tracks an interaction pattern for the advertisement. (EX-1003, ¶¶503-506.)

K. Claims 9, 17, 35, 43

As noted above (§VII.C.2), 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.” (EX-1009, 26:45-27:7.)

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.*” (EX-1009, 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.”).)

The Ramer-Cross combination therefore 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.*” (EX-1003, ¶¶507-10.)

L. Claims 10, 25, 36, 51

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).

(EX-1010, 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.*”

(EX-1010, 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”).)

Ramer similarly discloses that the speech patterns are detected and saved to assist in future interpretation of utterances. (EX-1009, 21:33-39.)

The Ramer-Cross combination therefore discloses all elements in these claims. (EX-1003, ¶¶511-16.)

M. Claims 11, 26, 37, 52

Cross further discloses dynamically updating the “*dictionary and phrase tables*” based on “*prior probabilities derived from the history of the current dialog and the prior dialogs.*” For example, 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).” (EX-1010, 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. (EX-1003, ¶¶517-18.)

N. Claims 12, 38

Claims 12 and 18 cover the same limitation in [14F] above (§VII.C.2). For the reasons discussed above in §VII.C.2 the Ramer-Cross combination discloses claims 12 and 18.

O. Claims 13, 15, 39, 41

As discussed above (§VII.C.2), Ramer discloses “*predetermined events*” that are detected by the “*adaptive misrecognition engine*” that cause the conversational language processor to reinterpret the words or phrases. Ramer discloses a predetermined event of “*receiving one or more of a subsequent input proximate in time to the natural language utterance.*” Specifically, 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.” (EX-1009, 50:65-51:3.)

Here, a POSITA would understand that “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.”

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.” (EX-1009, 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.*” (EX-1003, ¶¶521-26.)

P. Claims 16 and 42

Limitations [16P]-[16B] and [42P]-[42B] are substantively identical to limitations [1C]-[1D]/ [27C]-[27D]. For the reasons discussed above in §VII.B.4, the Ramer-Cross combination discloses these limitations. (EX-1003, ¶¶527-28.)

VIII. DISCRETIONARY DENIAL IS NOT APPROPRIATE

A. 35 U.S.C. §314(a)

Pursuant to the Stewart Memorandum dated March 26, 2025, titled “Interim Process for PTAB Workload Management,” Petitioner will respond to any discretionary denial arguments PO may raise through the Board’s bifurcated briefing process.

Nevertheless, to simplify the *Fintiv* analysis, and following the precedential *Sotera* decision, Petitioner stipulates that, if IPR is instituted, Petitioner will not

pursue in the related district court proceeding any ground that Petitioner raised or reasonably could have raised against the challenged claims during the instituted IPR.

B. 35 U.S.C. §325(d)

Advanced Bionics Part 1 is not satisfied because none of the prior art or arguments discussed herein is the same or substantially the same as art previously was presented during prosecution or in IPR2020-01390. The Board need not reach Part 2.

Moreover, the Petition's evidence warrants consideration because it presents a compelling case of unpatentability never considered by the Office. These circumstances weigh heavily against §325(d) discretionary denial.

IX. MANDATORY NOTICES

A. Real Party In Interest

The real parties-in-interest are Samsung Electronics Co. Ltd. and Samsung Electronics America, Inc.

B. Related Matters

To the best of Samsung's knowledge, the following is a list of judicial or administrative matter that would affect, or be affected by, a decision in the proceeding:

- *VB Assets, LLC v. Samsung Electronics Co. Ltd. et al.*, No. 2:24-cv-00828-JRG-RSP (E.D. Tex., filed Oct. 9, 2024)

- *VB Assets, LLC v. Amazon.com Servs. LLC*, No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019)
- *VB Assets, LLC v. Apple, Inc.*, No. 1:24-cv-01368-MN (D. Del., filed Dec. 13, 2024)
- *VB Assets, LLC v. SoundHound AI, Inc.*, No. 1:24-cv-01279-MN (D. Del., asserted via amended complaint Jan. 30, 2025)

C. Notice of Counsel and Service Information

Pursuant to 37 C.F.R. §§ 42.8(b)(3), 42.8, Petitioners designate the following lead and backup counsel:

Lead Counsel	Back Up Counsel
Lori Gordon (Reg. No. 50,633) GOODWIN PROCTER LLP 1900 N St. N.W. Washington, D.C. 20036 Phone: (202) 346-4000 Fax: (202) 346-4444 Gordon-ptab@goodwinlaw.com	Doug Kline (Reg No. 35,574) GOODWIN PROCTER LLP 100 Northern Avenue Boston, MA 02210 Tel: (617) 570 1000 Fax: (617) 523-1231 dkline@goodwinlaw.com
	Srikanth Reddy (<i>pro hac vice</i> application to be filed) GOODWIN PROCTER LLP 100 Northern Avenue Boston, MA 02210 Tel: (617) 570 1000 Fax: (617) 523-1231 sreddy@goodwinlaw.com
	Brian T. Drummond (Reg. No. 68,414) GOODWIN PROCTER LLP 100 Northern Avenue Boston, MA 02210 Tel: (617) 570 1000 Fax: (617) 523-1231 bdrummond@goodwinlaw.com
	Theodoros Konstantakopoulos (Reg. No. 74,155) GOODWIN PROCTER LLP 620 Eighth Avenue New York, NY 10018 Phone: (212) 813-8800 Fax: (212) 355-3333 tkonstantakopoulos@goodwinlaw.com

Petitioners consent to electronic service at the email addresses listed in the table above.

X. CONCLUSION

IPR is respectfully requested.

Date: April 16, 2025

Respectfully submitted,

/Lori A. Gordon/

Lori A. Gordon (Reg. No. 50,633)
Goodwin Procter LLP
1900 N Street, N.W.
Washington, D.C. 20036
Phone: (202) 346 4435
Gordon-ptab@goodwinlaw.com

Lead Counsel for Petitioner

CERTIFICATE OF WORD COUNT UNDER 37 CFR §42.24(d)

Pursuant to 37 C.F.R. §42.24(a), Samsung hereby certifies that portions of the above-captioned Petition for *Inter Partes* Review of U.S. Patent No. 7,818,176 in accordance with and reliance on the word count provided by the word-processing system used to prepare this Petition, that the number of words in this paper is 13,896. Pursuant to 37 C.F.R. §42.24(a), this word count is in compliance and excludes the table of contents, table of authorities, mandatory notices under §42.8, certificate of service, certificate of word count, appendix of exhibits, and any claim listing. This word count was prepared using Microsoft Word.

Date: April 16, 2025

Respectfully submitted,

/Lori A. Gordon/

Lori A. Gordon (Reg. No. 50,633)
Goodwin Procter LLP
1900 N Street, N.W.
Washington, D.C. 20036
Phone: (202) 346 4435
Gordon-ptab@goodwinlaw.com

Lead Counsel for Petitioner

CERTIFICATE OF SERVICE

The undersigned certifies that a true copy of the Petition for *Inter Partes* Review of U.S. Patent No. 7,818,176 together with all exhibits identified in the above Table of Exhibits and Petitioners' Power of Attorney, have been served on the Patent Owner via FedEx Next Business Day Delivery on the below date, at the correspondence address of record as listed on the Patent Center:

Sheppard Mullin Richter & Hampton LLP
650 Town Center Drive
10th Floor
Costa Mesa, CA
Attorneys for VB Assets, LLC

Courtesy copies of the foregoing document were served by e-mail on the following counsel of record for Patent Owner in the concurrent litigation:

Theodore Stevenson, III (TX 19196650)
Brady Randall Cox (TX 24074084)
ALSTON & BIRD LLP
2200 Ross Avenue, Suite 2300
Dallas, TX 75201
Telephone: (214) 922-3400
Facsimile: (214) 922-3899
ted.stevenson@alston.com
brady.cox@alston.com

Erik John Carlson (CA 265167)
Caleb J. Bean (CA 299751)
Srishti Ghosh (CA 354393)
ALSTON & BIRD LLP
350 South Grand Avenue, 51st Floor

Los Angeles, CA 90071
Telephone: (213) 576-1000
Facsimile: (213) 576-1100
erik.carlson@alston.com
caleb.bean@alston.com
maddie.ghosh@alston.com

Natalie C. Clayton (NY 4409538)
ALSTON & BIRD LLP
90 Park Avenue, 15th Floor
New York, NY 10016
Telephone: (212) 210-9400
Facsimile: (212) 210-9444
natalie.clayton@alston.com

David Greenbaum (NY 2947455)
GREENBAUM LAW LLC
210 Allison Court
Englewood, NJ 07631
david@greenbaum.law
Attorneys for VB Assets LLC

Date: April 16, 2025

/Lori A. Gordon/

Lori A. Gordon (Reg. No. 50,633)
Goodwin Procter LLP
1900 N Street, N.W.
Washington, D.C. 20036
Phone: (202) 346 4435
Gordon-ptab@goodwinlaw.com

Lead Counsel for Petitioner

APPENDIX A – LIST OF CLAIMS

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

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

[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:

[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

[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;

[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;

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

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

[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.

[3P] The method of claim 2, further comprising:

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

[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.

[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.

[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.

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

[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;

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

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

[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.

[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.

[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.

[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.

[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.

[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.

[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.

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

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

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

[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;

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

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

[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.

[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.

[16P] The method of claim 14, wherein recognizing the words or phrases in the natural language utterance includes:

[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

[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.

[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.

[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.

[19P] The method of claim 18, further comprising:

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

[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.

[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.

[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.

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

[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;

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

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

[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.

[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.

[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.

[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.

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

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

[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:

[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

[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;

[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;

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

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

[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.

[29P] The system of claim 28, wherein the conversational language processor is further configured to:

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

[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.

[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.

[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.

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

[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

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

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

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

[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;

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

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

[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.

[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.

[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:

[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

[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.

[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.

[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.

[45P] The system of claim 44, wherein the conversational language processor is further configured to:

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

[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.

[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.

[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.

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

[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

[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.

[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.

[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.

[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.

[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.