

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

IPR2025-00870

U.S. Patent No. 10,755,699

**PETITION FOR *INTER PARTES* REVIEW OF
U.S. PATENT NO. 10,755,699**

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 Challenge1

 A. Prior Art.....1

 B. Grounds For Challenge2

III. The '699 Patent.....2

 A. Patent Overview2

 B. Level of Ordinary Skill3

 C. Claim Construction.....4

IV. GROUND 1: SmartKom-Kobsa Combination Renders Claims 1-22
Obvious.....4

 A. Combination Overview4

 1. SmartKom4

 2. Kobsa.....6

 3. Motivation to Combine7

 B. Independent Claims 1, 12.....9

 1. Preambles [1P.1]-[1P.2]/[12P] and Limitation [12A]9

 a) System and Method for “Generating Natural
Language System Responses”9

 b) “Computer System”11

 2. Receiving A User Input [1A]/[12B]12

 3. Recognizing Words/Phrases [1B]/[12C]13

 4. Identifying Context [1C]/[12D]15

5.	Determining An Interpretation [1D]/[12E].....	23
6.	Short-Term and Long-Term Knowledge Limitations.....	27
a)	Accumulating Short-Term Knowledge [1E.1]- [1E.2]/ [12F.1]-[12F.2].....	29
b)	Accumulating Long-Term Knowledge [1F]/[12G].....	34
7.	Identifying Manner Of Speaking [1G]/[12H].....	37
8.	Generating A Response [1H]/[12I].....	41
C.	Dependent Claims	46
1.	Claims 2, 13	46
2.	Claims 3, 9, 14, 20	48
3.	Claims 4, 15	50
4.	Claims 5, 16	54
5.	Claims 6, 17	56
6.	Claims 7, 18	57
7.	Claims 8, 10, 19, 21	58
8.	Claims 11, 22	60
V.	GROUND 2: Barbara-Ross-Kellner Combination Renders Obvious Claims 1-22.....	61
A.	Combination Overview	61
1.	Barbara	61
2.	Ross	63
3.	Kellner.....	64
4.	Motivation to Combine Barbara With Ross	65
5.	Motivation To Combine Barbara And Ross With Kellner	68

B.	Independent Claims	70
1.	Preambles [1P.1]/[1P.2]/[12P]/[12A]	70
2.	Receiving An Utterance [1A]/[12B]	72
3.	Recognizing Words or Phrases [1B]/[12C]	73
4.	Identifying a Context [1C]/[12D]	74
5.	Determining an Interpretation [1D]/[12E]	74
6.	Short-Term Knowledge [1E.1]/[1E.2]/[12F.1]/[12F.2]	75
7.	Long-Term Knowledge [1F]/[12G]	77
8.	Identifying a Manner [1G]/[12H]	78
9.	Generating A Response [1H]/[12I]	80
C.	Dependent Claims	82
1.	Claims 2, 13	82
2.	Claims 3, 14	83
3.	Claims 4, 15	83
a)	Contextual Signifiers And/Or Grammatical Rules [4A]/[15A]	83
b)	Response Based On Contextual Signifiers And/Or Grammatical Rules [4B]/[15B]	84
4.	Claims 5, 16	85
5.	Claims 6, 17	86
6.	Claims 7, 18	87
7.	Claims 8, 19	89
8.	Claims 9, 20	90
9.	Claims 10, 21	91

10. Claims 11, 2291

VI. Stipulation.....94

VII. Mandatory Notices.....94

 A. Real Party In Interest.....94

 B. Related Matters.....94

 C. Notice of Counsel and Service Information.....95

VIII. Conclusion96

EXHIBIT LIST

Exh.	Reference
1001	U.S. Patent 10,755,699 to Baldwin et al. (“the ’699 patent”)
1002	Prosecution History for U.S. Patent 10,755,699
1003	Declaration of Stuart Lipoff in Support of <i>Inter Partes</i> Review of U.S. Patent 10,755,699
1004	Curriculum Vitae of Stuart Lipoff
1005	<i>SmartKom: Foundations of Multimodal Dialogue Systems</i> , Springer Science + Business Media, Inc. (Wolfgang Wahlster ed., 2006) (“SmartKom”)
1006	<i>User Models in Dialog Systems</i> , Springer-Verlag (Kobsa & Wahlster eds., 1989) (“Kobsa”)
1007	U.S. Publication 2004/0101198 to Barbara (“Barbara”)
1008	U.S. Publication 2002/0173960 to Ross et al. (“Ross”)
1009	Declaration of June Munford in Support of <i>Inter Partes</i> Review of U.S. Patent 10,755,699
1010	<i>Amazon.com et al. v. VB Assets LLC</i> , IPR2020-01367, Paper 27 (PTAB March 7, 2022).
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)
1012	Susann Luperfoy, <i>Discourse PEGS: A Computational Analysis of Context-Dependent Referring Expressions</i> (Dec. 1991) (Ph.D. Dissertation, University of Texas at Austin) (“Luperfoy”)
1013	<i>Microsoft Computer Dictionary</i> , Microsoft Press (2002)

Exh.	Reference
1014	Robert Porzel & Iryna Gurevych, <i>Contextual Coherence in Natural Language Processing</i> , 2680 Lecture Notes in Computer Science 272 (“Porzel”)
1015	Reserved
1016	Niels Ole Bernsen et al., <i>Designing Interactive Speech Systems: From First Ideas to User Testing</i> , Springer (1998) (“Bernsen”)
1017	H.P. Grice, <i>Logic and Conversation in 3</i> Syntax and Semantics 41, Academic Press (1975) (“Grice”)
1018	Xuedong Huang et al., <i>Spoken Language Processing</i> , Prentice Hall PTR (2001) (“Huang”)
1019	Lea Krause & Piek Vossen, <i>The Gricean Maxims in NLP – A Survey</i> , Proceedings of the 17th International Natural Language Generation Conference (2024) (“Krause”)
1020	Michael F. McTear, <i>Spoken Dialogue Technology: Toward the Conversational User Interface</i> , Springer (2002)
1021	Sebastian Möller, <i>Quality of Telephone-Based Spoken Dialogue Systems</i> , Springer Science + Business Media (2005) (“Möller”)
1022	Susan E. Brennan, <i>The Multimedia Articulation of Answers in a Natural Language Database Query System</i> , Second Conference on Applied Natural Language Processing, (1988) (“Brennan”)
1023	U.S. Publication 2002/0065651 to Kellner (“Kellner”)

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

(“Petitioners”) petition for *inter partes* review of claims 1-22 of U.S. Patent 10,755,699 (EX-1001).

I. Grounds for Standing

Petitioners certify the ’699 patent is available for IPR; Petitioners are not barred or estopped.

II. Identification of Challenge

A. Prior Art

The ’699 patent claims priority through a series of continuations/divisionals to U.S. Patent 8,073,681 (“’681 patent”), filed October 16, 2006. Petitioners do not acquiesce that the claims are entitled to the priority date. Regardless, each applied reference was filed or published before this date.

1. **“*SmartKom: Foundations of Multimodal Dialogue Systems,*”** (“SmartKom”; EX-1005) is prior art under 35 U.S.C. §102(a) because it was publicly available by August 4, 2006. (EX-1005, ¶¶6-9.)
2. **“*User Models in Dialog Systems,*” Kobsa & Wahlster eds.** (“Kobsa”; EX-1006) is prior art under 35 U.S.C. §102(b) because it was publicly available by December 5, 1989. (EX-1006, ¶¶10-13.)
3. **U.S. Publication 2004/0101198 to Barbara** (“Barbara”; EX-1007), published May 27, 2004, is prior art under 35 U.S.C. §102(b).

4. **U.S. Publication 2002/0173960 to Ross** (“Ross”; EX-1008), published November 21, 2002, is prior art under 35 U.S.C. §102(b).

5. **U.S. Publication 2002/0065651 to Kellner** (“Kellner”; EX-1023), published May 30, 2002, is prior art under 35 U.S.C. §102(b).

B. Grounds For Challenge

Ground		Claims	Prior Art
1	103	1-22	SmartKom+Kobsa
2	103	1-22	Barbara+Ross+Kellner

III. The '699 Patent

A. Patent Overview

The '699 patent discloses “a cooperative conversational voice user interface.” (Ex-1001, 7:33-35.) The system receives an utterance (input 105), which is processed by speech recognition engine 110 to generate preliminary interpretations. (Ex-1001, 7:50-54, Figure 1(below).) The preliminary interpretations are provided to conversational speech engine 115 for further processing. (Ex-1001, 7:63-66). Conversational speech engine 115 communicates with databases 130 to generate a response. (EX-1001, 8:2-5.)

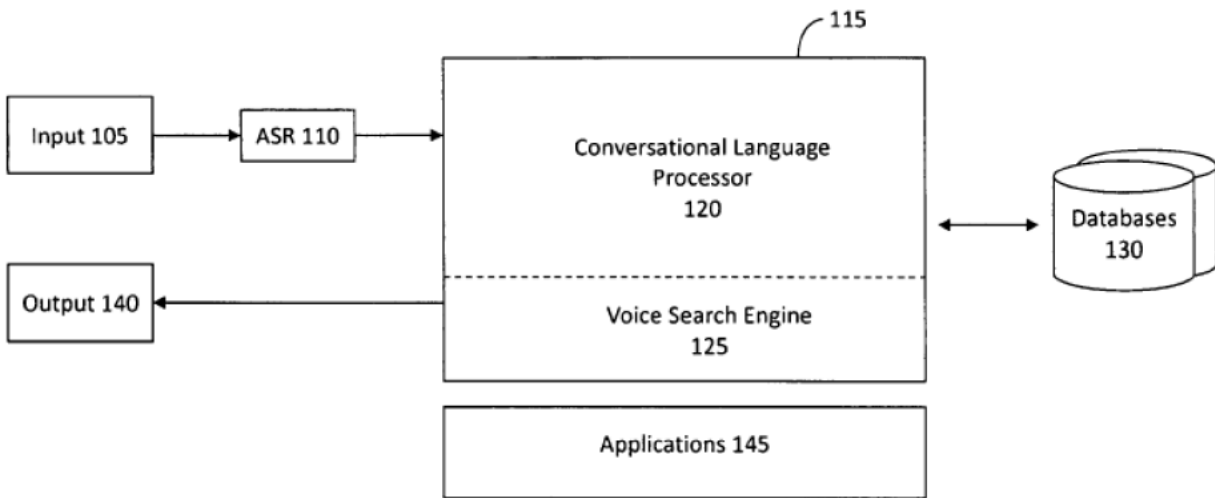


Figure 1

'699 Patent, Figure 1

B. 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.¹

¹ The Board adopted this definition, which Patent Owner (“PO”) did not dispute, in IPR2020-01367, challenging the ’681 patent. (EX-1010, 8.)

C. Claim Construction

In *VB Assets, LLC v. Amazon.com Servs. LLC*, 1:19-cv-01410-MN, the parties agreed the term “speech recognition engine” recited in the ’681 patent means “software or hardware that recognizes the words or phrases in the natural language utterance.” (EX-1011, 2.) The term “speech recognition engine” also appears in claims [7]/[18] of the ’699 Patent. Petitioners applied this construction.

Petitioners believe no other term requires construction to resolve patentability in this proceeding.

IV. GROUND 1: SmartKom-Kobsa Combination Renders Claims 1-22² Obvious.

A. Combination Overview

1. SmartKom

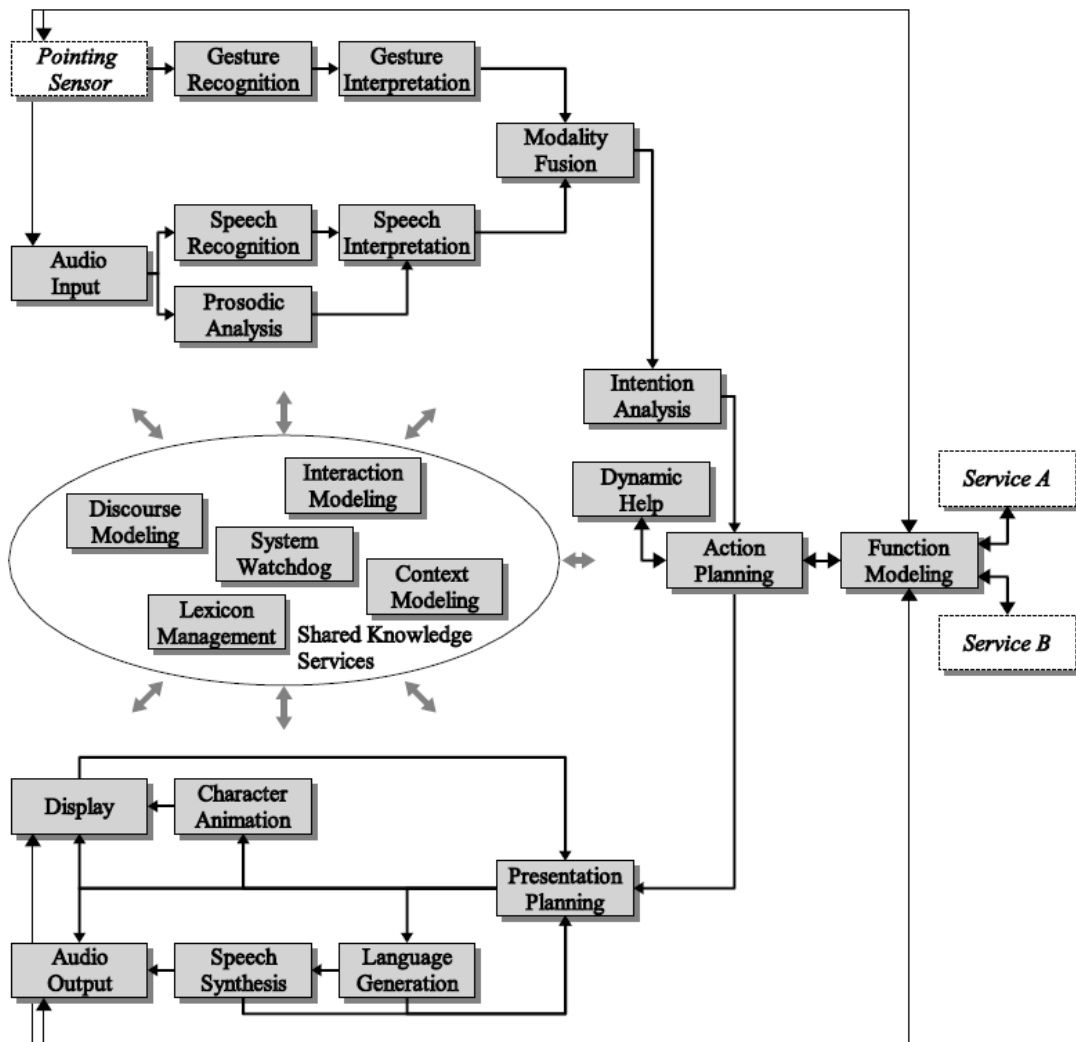
SmartKom “provides a comprehensive overview” of the SmartKom system and research results developed as part of the SmartKom initiative, begun in spring 1999. (EX-1005, VI, 31-32; EX-1003, ¶¶104-122 (providing context for end-end system).) The SmartKom system “provides full symmetric multimodality in a mixed-initiative dialogue system with an embodied conversational agent.” (EX-1005, 3.) The generic software architecture provides a set of functional components, each represent a “processing unit[] of the executing system.” (EX-

² Listing of challenged claims provided in the Appendix.

1005, 59, Figure 2 (60)³(below).) SmartKom describes three exemplary application scenarios SmartKom-Public (EX-1005, 63-64), SmartKom-Home (EX-1005, 65-66) and SmartKom-Mobile (EX-1005, 66-67)) derived from this generic architecture. The shaded components indicate components “reused in all application cases.” (EX-1005, 59.)

³ Because SmartKom is a series of papers, Figure numbers are reused.

Petitioner provides the page number in the Figure label.



SmartKom, Figure 2(60)

2. Kobsa

Kobsa “provide[s] a rather coherent survey of the field of user modeling.”

(EX-1006, V.) Kobsa describes “user models” which are each a knowledge source “contain[ing] explicit assumptions on all aspects of the user that may be relevant to the dialog behavior of the system.” (EX-1006, 6.) Kobsa presents several exemplary user models storing user information (e.g., user domain expertise and/or

preferences) and describes user modeling components which “incrementally construct a user model,” “store, update and delete entries” and “supply other components of the system with assumptions about the user.” (EX-1006, 6.)

3. Motivation to Combine

A POSITA would have been motivated to combine Kobsa’s teachings regarding user models and user modeling with SmartKom’s dialogue system. (EX-1003, ¶¶126-131.) Kobsa is in the same field as the SmartKom and the ’699 patent—speech recognition systems. (EX-1001, 1:28-29 (“The invention relates to a cooperative conversational model for a human to machine voice user interface”); EX-1005, 3 (SmartKom “represents a new generation of multimodal dialogue systems”); EX-1006, 4 (describing user models “in natural-language dialog systems”); EX-1003, ¶126.)

SmartKom explicitly motivates the combination, mentioning use of a user model and/or user profiles/preferences in its context modeling and presentation planning. (*See, e.g.*, EX-1005, 16 (use of user model by the presentation planner), 274 (use of user model in context modeling), 323 (referencing use of “UserKnowledge”), 407 (“Given an appropriate user model, personal preferences ... can be used.”); EX-1003, ¶127.) SmartKom also mentions storing preferences stated by a user during a conversation. (*See* EX-1005, 336.) However, SmartKom provides limited additional details regarding the content of the user models and

techniques for user modeling. Accordingly, based on the suggestions in SmartKom, a POSITA would have been motivated to search for references that describe user models and would have been led to Kobsa which provides a “coherent survey” of the topic. (EX-1003, ¶128; EX-1006, V.) Kobsa is co-edited by Wolfgang Wahlster who edited SmartKom and is Scientific Director of the SmartKom project, further leading a POSITA to Kobsa for further details of user models. (*Id.*)

Kobsa also motivates the combination describing benefits of user models for “interact[ing] with people in an intelligent and cooperative manner.” (EX-1006, 411, EX-1003, ¶129.) Kobsa stresses that user models allow the system to “tailor[] object descriptions to the user’s level of expertise” and “adapt[] an expert system’s response behavior to the background knowledge of its users.” (EX-1006, 196, 416.) A POSITA would have therefore been motivated to add Kobsa’s teachings into SmartKom to improve the user experience by providing user-specific and tailored responses. (EX-1003, ¶129.)

The combination is also nothing more than use of a known technique (Kobsa’s user models and user modeling) to improve similar devices (SmartKom’s dialogue system) in the same way (providing any additional knowledge source of interpretation and response generation). *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007); EX-1003, ¶130.

A POSITA would have had a reasonable expectation of success and the results of the combination would have been predictable because SmartKom uses a standard software architecture and standard storage structures. (See EX-1005, Figure 2(60); EX-1003, ¶131.) Kobsa’s user model and user modeling components are also merely storage and software constructs. A POSITA would have therefore been able to implement Kobsa’s user models/modeling based on the teachings of SmartKom with predictable results. (EX-1003, ¶131.)

B. Independent Claims 1, 12

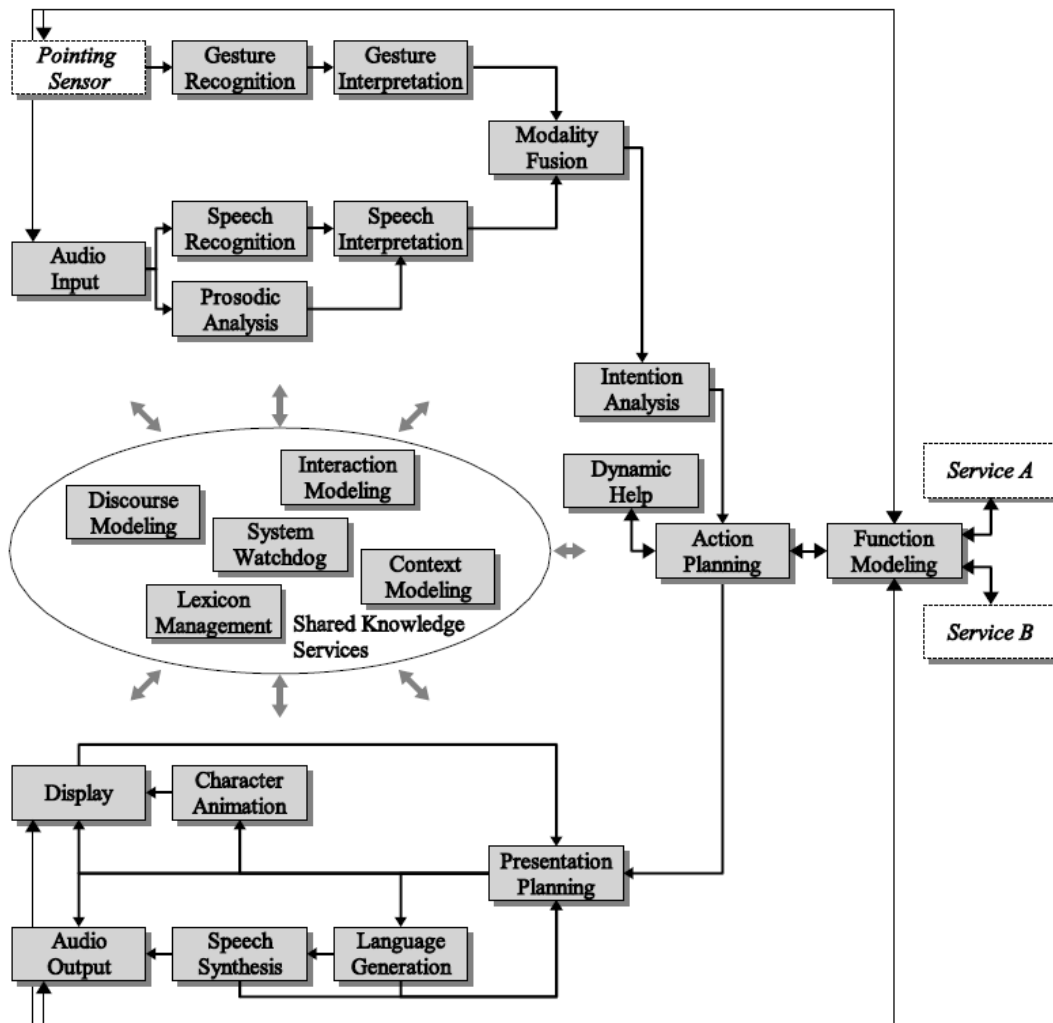
1. Preambles [1P.1]-[1P.2]/[12P] and Limitation [12A]

a) System and Method for “Generating Natural Language System Responses”

SmartKom discloses a “*system*”⁴ [12P], the SmartKom system which “represents a new generation of multimodal dialogue systems.” (EX-1005, 3.)

The generic software architecture of the SmartKom system is illustrated in Figure 2(60) below. (EX-1005, 60; *see also*, EX-1005, 275.) This software provides a “*method*” [1P] performing the claim 1 actions.

⁴ Claim language indicated by italics throughout.



SmartKom, Figure 2(60)

The SmartKom system and method “*generat[e] natural language system responses.*” In SmartKom, the “natural language generation (NLG) module” generates “not just plain text but rather generates complete syntactic structure and discourse information to supply highly annotated information structures for the concept-to-speech approach of SmartKom” for speech synthesis. (EX-1005, 401.) And the speech synthesis module “not only produces audio data but also produces

a detailed representation of the phonemes and their exact time points (in milliseconds) inside the audio signal.” (EX-1005, 396.) This approach allows the character animation module to “generate[] a lip animation script for Smartakus [display character] that is then executed during speech output.” (*Id.*)

In the SmartKom, the “*natural language system responses [are] adapted based on a user's manner of speaking.*” (See also, §§IV.B.7-8.) SmartKom’s “[i]nteraction modeling is concerned with different aspects like available modalities and user preferences for specific forms of communication as well as the affective state of the user.” (EX-1005, 61.) SmartKom’s “interaction model allows one to dynamically adapt the communicative behaviour of the system.” (*Id.*) For example, the model “describes the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete sentences and average length of input.” (EX-1005, 323.) All of these “help to adapt the **language generation** in order to reflect the user’s style.” (*Id.*) Kobsa also stresses that user models allow the system to “tailor[] object descriptions to the user’s level of expertise” and “adapt[] an expert system’s response behavior to the background knowledge of its users.” (EX-1006, 196.)

b) “Computer System”

SmartKom discloses “*a computer system*” [1P.2] including “*one or more physical processors executing one or more computer program instructions which,*

when executed” [1P.2]/[12A] “*perform the method*”[1P.2] or “*configure the one or more physical processors*” [12A].

SmartKom describes implementations of its software that execute on a server with at least one physical processor. (See EX-1005, 13, 440.) SmartKom’s software, therefore, provides a “*computer-implemented method*” [1P.1]. Although not explicit in SmartKom, a POSITA would have known, based on general knowledge in the field, that SmartKom’s software components include “*computer program instructions*”⁵ executed by “*one or more physical processors*” [1P.2]/[12A]. (EX-1003, ¶¶133-135, *citing* EX-1013, 489 (software is “Computer program; instructions that make hardware work”).)

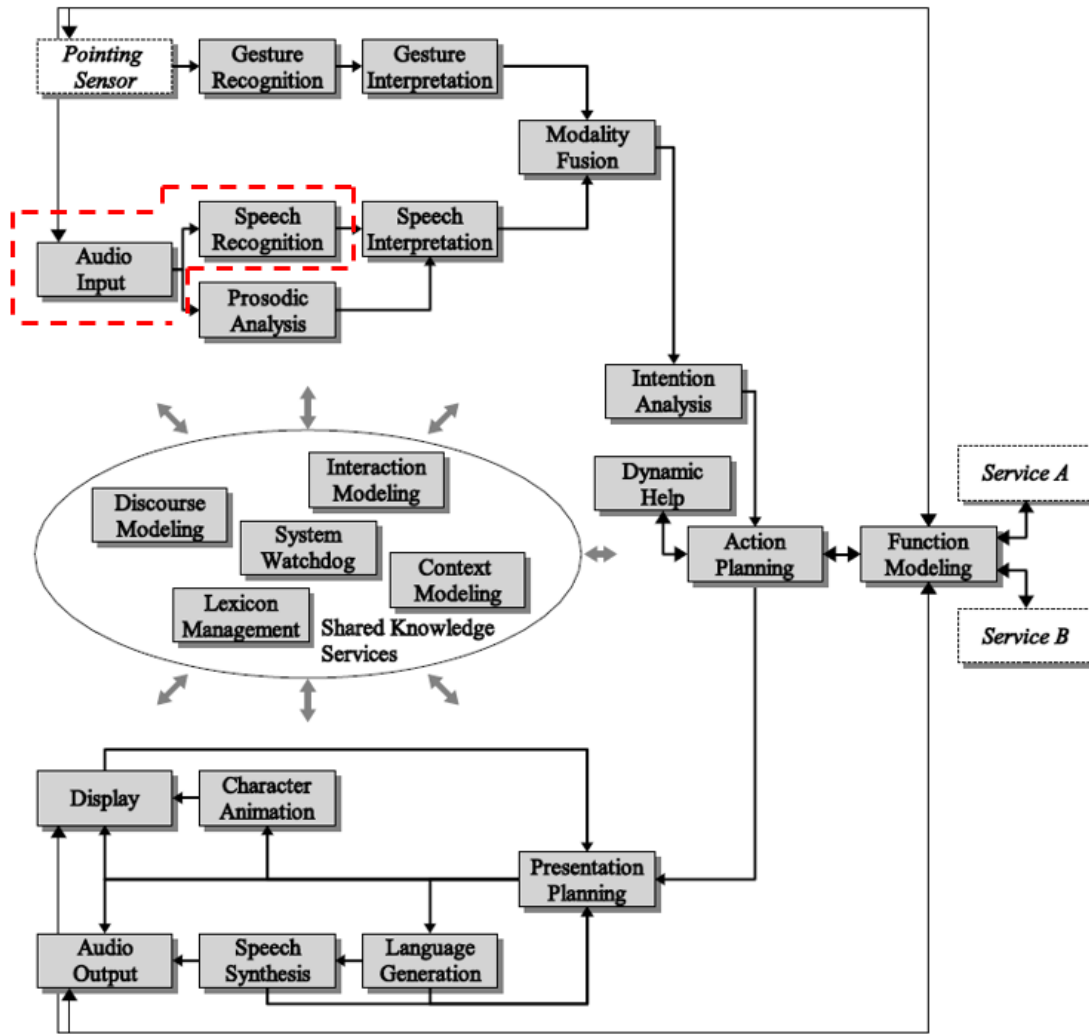
2. Receiving A User Input [1A]/[12B]

SmartKom discloses a “*computer system*”/“*physical processor[s]*” that receive “*a user input comprising a natural language utterance*” [1A]/[12B].

The SmartKom system includes an audio input (e.g., a microphone) that receives voice (speech) input from a user and provides an acoustic input signal to a speech recognition component that “*transforms the acoustic input signal.*” (See

⁵ The ’699 patent does not use the term “*instructions.*” At most, it mentions a “*conversational language processor.*” (See EX-1001, 8:23-28.)

EX-1005, 85, 440, 457, Figure 2(60).) The audio input therefore “receive[s] a user input comprising a natural language utterance.”

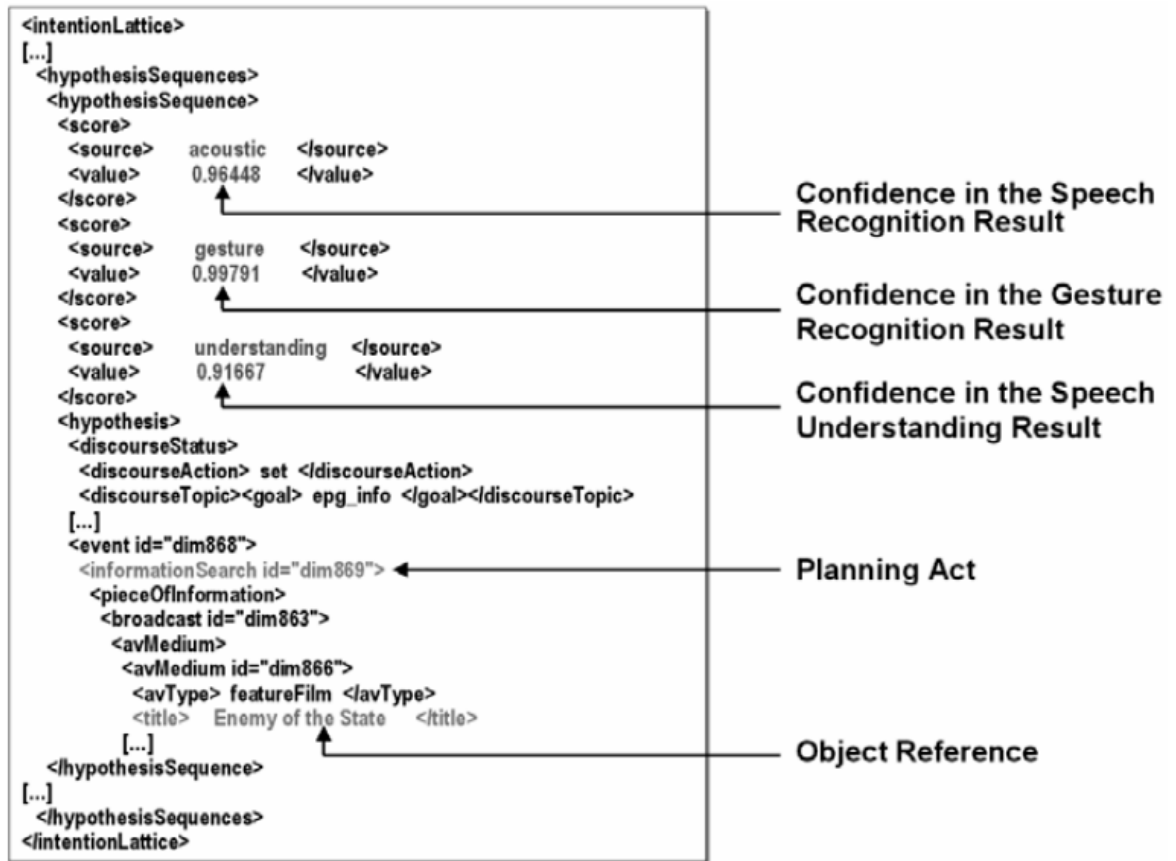


SmartKom, Figure 2(60)

3. Recognizing Words/Phrases [1B]/[12C]

SmartKom discloses a “computer system”/“physical processor[s]” that recognizes “one or more words or phrases from the natural language utterance.”

The SmartKom system includes a “**speech recognition engine**” that “transforms the acoustic signal” representing the user’s vocal utterance “into a **sequence of hypothesized words** in orthographic representation.” (EX-1005, 86, Figure 2(60).) That is the speech recognition engine “*recogniz[es] one or more words or phrases from the [received] natural language utterance.*” SmartKom’s system also uses “**confidence measures**” to “estimate confidences for the correctness of the words hypothesized by the recognizer.” (EX-1005, 85; *see also* EX-1005, 96 (“confidence of each word is estimated as the posterior probability for the word being correct”).) SmartKom notes that such “confidence measures” were “common practice in state-of-the-art speech recognition systems.” (EX-1005, 96.) The speech recognizer provides the hypotheses to the natural language processing (speech interpretation) module which “transform[s] the word lattice sent by the speech recognizer into a list of hypotheses representing ... possible user intentions.” (EX-1005, 195.) These hypotheses are provided, along with a speech recognizer confidence score and speech understanding score, to the intention recognizer in a data structure referred to as the intention lattice, as shown in the exemplary fragment below. (*See* EX-1005, Figure 8(15), 204.)



SmartKom, Figure 8(15)

4. Identifying Context [1C]/[12D]

The SmartKom-Kobsa combination discloses a “*computer system*”/ “*physical processor[s]*” that identifies “*a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance*” [1C]/[12D]. (EX-1003, ¶¶136-147.) SmartKom notes that “[s]peakers may not always be aware of the potential ambiguities inherent in their utterances” and “leave it to the context to disambiguate and specify the message.”

(EX-1005, 275.) Specifically, “to interpret the utterance correctly, the addressee must employ several context-dependent resources.” (EX-1005, 275.)

In SmartKom, the intention recognizer employs “context-dependent resources” by engaging the discourse modeler and context modeler which work together “to enrich the information in the hypotheses with context knowledge.” (EX-1005, 287.) The “enrichments are evaluated by the discourse model and the context model to rate the quality of the augmentation of a hypothesis with knowledge from the discourse and the surrounding world.” (EX-1005, 287.)

As part of this process, “the discourse modeler receives a set of hypotheses” and enriches each “with previous discourse information” (discourse state; “*short-term knowledge*” (§IV.B.5)) and context-domain information from the domain model (“*long-term knowledge*” (§IV.B.5)). (EX-1005, 20, 238 (“each new user contribution has to be interpreted in the light of the previous discourse context”), 240 (discourse modeler “interprets [hypotheses] with respect to the current discourse context”).) Specifically, the discourse modeler associates domain-context information with the utterance through a typed feature structure (TFS). (EX-1005, 256.)

The following example illustrates this context identification. Here, “the user has requested information about tonight’s television program” (utterance-1). (EX-

1005, 262-263.) The TFS representation of utterance-1 (part of the discourse history in the discourse state) is shown in Figure 2(264).

U-1: *What’s on TV tonight?*

SYS: (Displays the TV program listings) *Here you see a list of movies running tonight.*

U-2: *Is there a movie with Nicole Kidman?*

(EX-1005, 263.)

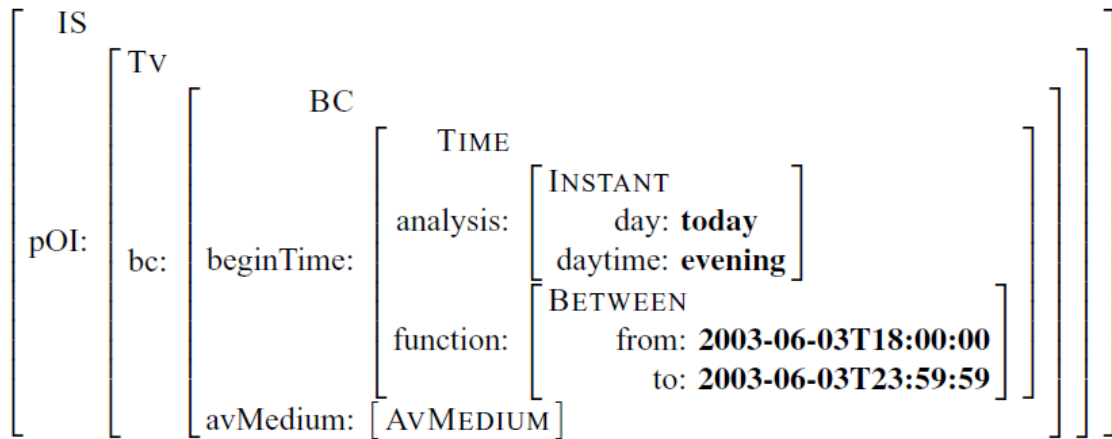


Fig. 2. TFS representation for the utterance *What’s on TV tonight?* (*IS* = InformationSearch, *pOI* = pieceOfInformation, *Tv* = TvProgram, *bc* = broadcast)

SmartKom, Figure 2(264)

The user then “wants to reduce the number of displayed movies by uttering *Is there a movie with Nicole Kidman?*” (utterance-2). (EX-1005, 263.) “This user request, if interpreted separately, leads to at least two intention hypotheses— one where the user requests information about the current TV program” (U-2-hypoB)

and “one where the user requests information about the current cinema program” (U-2-hypoA). (EX-1005, 263.) The TFS for the first interpretation is shown in Figure 3(264). As shown, U-2-hypoB is associated with the context-domain of a cinema program. (EX-1005, 264.) The TFS for the second interpretation is shown in Figure 4(264). As shown, U-2-hypoA is associated with the context-domain of a broadcast TV performance. (EX-1005, 264.) That is, the discourse modeler enriches each with context-domain information—i.e., identifies a context.

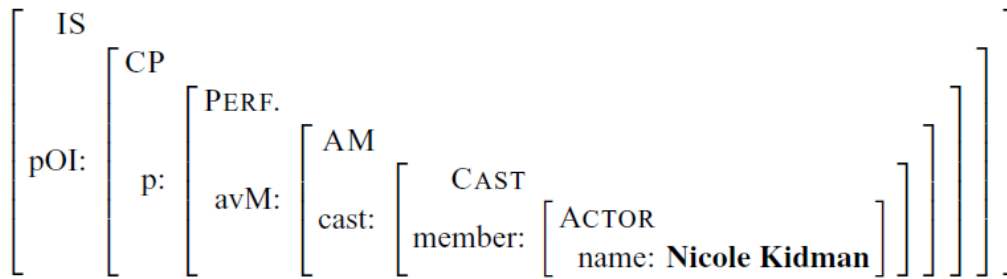


Fig. 3. One possible interpretation of the utterance *Is there a movie with Nicole Kidman?* (*IS* = InformationSearch, *pOI* = pieceOfInformation, *CP* = CinemaProgram, *p* = performance, *P* = Performance, *bc* = Broadcast, *Bc* = Broadcast)

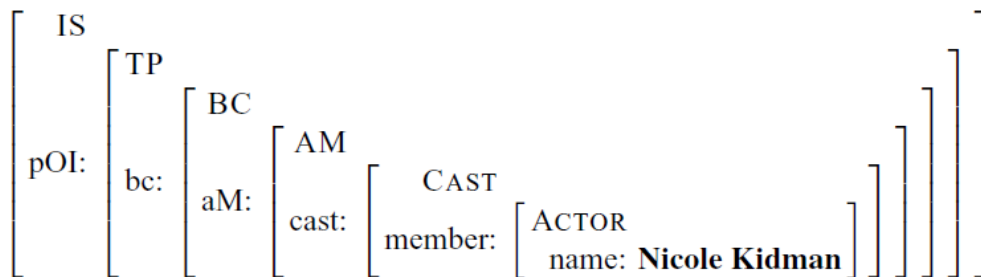


Fig. 4. A second possible interpretation of the utterance *Is there a movie with Nicole Kidman?* (*IS* = InformationSearch, *pOI* = pieceOfInformation, *CP* = CinemaProgram, *p* = performance, *P* = Performance, *aM* = avMedium, *AM* = AvMedium)

SmartKom, Figures 3-4(264)

Each of these hypotheses, enriched with context-domain information, is then overlay[ed] with the prior utterance (U-1) and scored. (EX-1005, 266.) Because U-1 is associated with TV context domain, the TV intention hypothesis U-2-hypoB is more likely and receives a higher score. (*Id.*) The interpretation/enrichment of U-2 with respect to the previous discourse context is shown in Figure 5(265). As

shown, the hypothesis has been enriched with the identified context based on the discourse history and context-domain (broadcast-TV). (EX-1003, ¶141.)

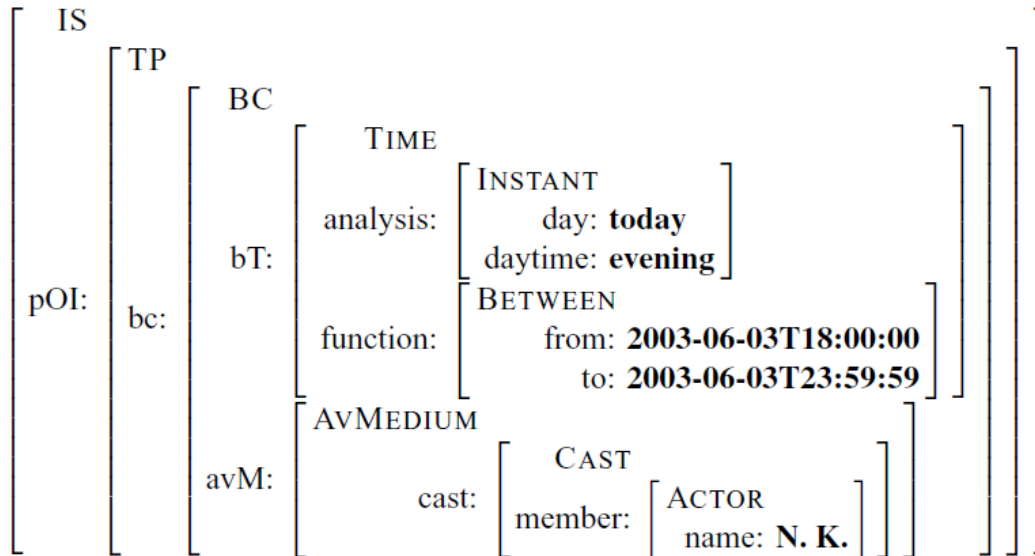


Fig. 5. Interpretation of the utterance *Is there a movie with Nicole Kidman?* with respect to the previous discourse context

SmartKom, Figure 5(265)

In addition to the above enrichment through discourse state and context-domain, the hypotheses are enriched with context from the context model. (EX-1005, 285, 287.) The SmartKom-Kobsa system uses this contextual knowledge for “**lexical and pragmatic disambiguation**, decontextualization [sic] of domain and common-sense knowledge that was left implicit by the user” and “for estimat[ion of] an overall coherence score that is used in intention recognition.” (EX-1005,

269.) To perform these actions, SmartKom “link[s] the **context model to interfaces providing contextual information.**” (EX-1005, 277-278.)

Specifically, the context model uses a set of contexts, illustrated in Table 1(275), associated with a knowledge store (referred to as a “model”). (EX-1005, 274-275.) The “*context[s]*” provided by the model are used to enrich the “intention lattice” hypotheses through the discourse/context modelers, as discussed below. (EX-1005, 305-306.)

Table 1. Contexts, content and knowledge sources

Types of context	Content	Knowledge store
Dialogical context	What has been said by whom	Dialogue model
Ontological Context	World/conceptual knowledge	Domain model
Situational context	Time, place, etc.	Situation model
Interlocutionary context	Properties of the interlocutors	User model

SmartKom, Table 1(274)

The context model provides “context-specific insertions” to enrich each hypothesis (e.g., “provide hitherto implicit knowledge concerning what is talked about”). (EX-1005, 277.) For example, the “emotional state of the user” (i.e., the user state) is added. (EX-1003, ¶146.) Situational context, including location and time of the utterance, is also added, as shown in exemplary Table 2(277) below. In the combination of SmartKom-Kobsa, a hypothesis is enriched with knowledge from the user model such as the user’s preferences or expertise in the identified

context-domain. (EX-1005, 274.) These enrichments assist with interpretation of user intent as well as shaping response generation. (§§IV.B.5, 7-8.)

Table 2. Context-specific insertions into a sample intention hypothesis resulting from the interpretation of a speech recognition hypothesis

<pre><informationSearchProcess> <entertainment> <performance> <cinema> <contact> <address> <town> here </town> </address> </contact> </cinema> <time> <beginTime> <at> now </at> </beginTime> </time> </performance> </entertainment> </informationSearchProcess></pre>	<pre><contact> <x> 70.345 </x> <y> 49.822 </y> <town> Heidelberg </town> </contact> <time> <at> 19:00:00T26:08:03 </at> </time></pre>
	<pre><scores> <contextualCoherence> 0.46 </contextualCoherence> </scores></pre>

SmartKom, Table 2(277)

Thus, the context modeler of the SmartKom-Kobsa combination identifies a context based on dynamic situational knowledge, dynamic instances (e.g., position of user when providing current utterance), and common ground knowledge, and user-specific profile information (e.g., experience, preferences, etc.) and incorporates this context with the discourse and context-domain context (e.g.,

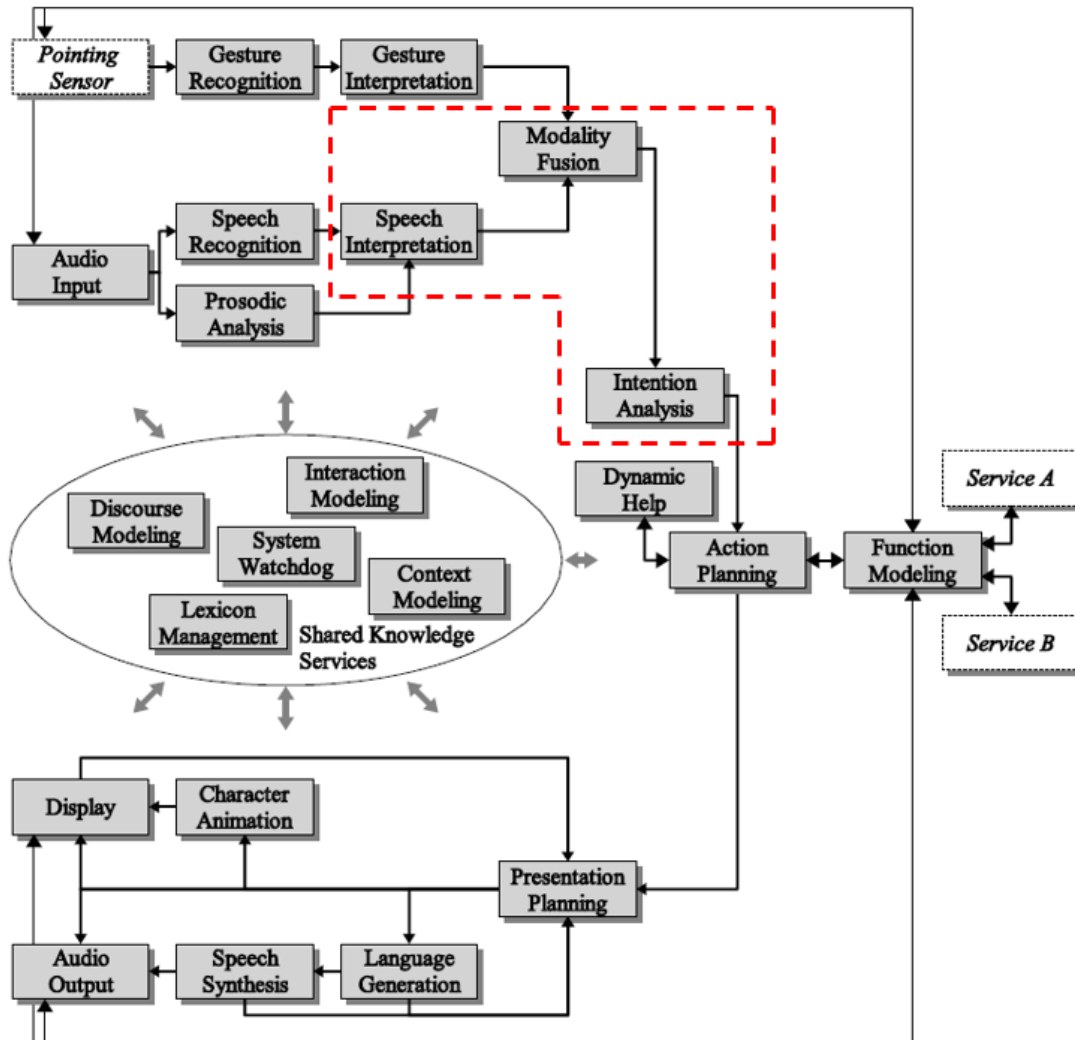
television, cinema program, etc.) in each hypothesis. (EX-1003, ¶147; *see, e.g.*, EX-1005, 364 (“SmartKom domains such as electronic program guides (EPG) for TV’s, cinema programs, and movie information”), 21.)

The context modeler also “compute[s] [a] contextual coherence score[.]” (EX-1005, 276, 271.) The scoring task involves using the “ontological domain context to measure the semantic coherence of the individual interpretation” and “using dynamic situational and discourse information, e.g., previous ontological contexts of prior turns.” (EX-1005, 271.) For example, a companion paper by Porzel, explains that “a panel of experts” including a “*situation expert* observing, e.g., time, data, proximity, accessibility”, “*user expert* observing, e.g., interests, transportation, thrift”, “*discourse expert* observing, e.g., referential status, discourse accessibility,” and an “*ontological expert* observing, e.g., object types and object roles” vote and these votes are combined to obtain the contextual coherence score. (EX-1014, 283 (emphasis in original).) Table 2(277) illustrates a contextual coherence score added to the intention lattice.

5. Determining An Interpretation [1D]/[12E]

SmartKom discloses a “*computer system*”/“*physical processor[s]*” that determines “*an interpretation of the natural language utterance based on the identified context*” [1D]/[12E]. SmartKom’s intention recognizer “has the task to

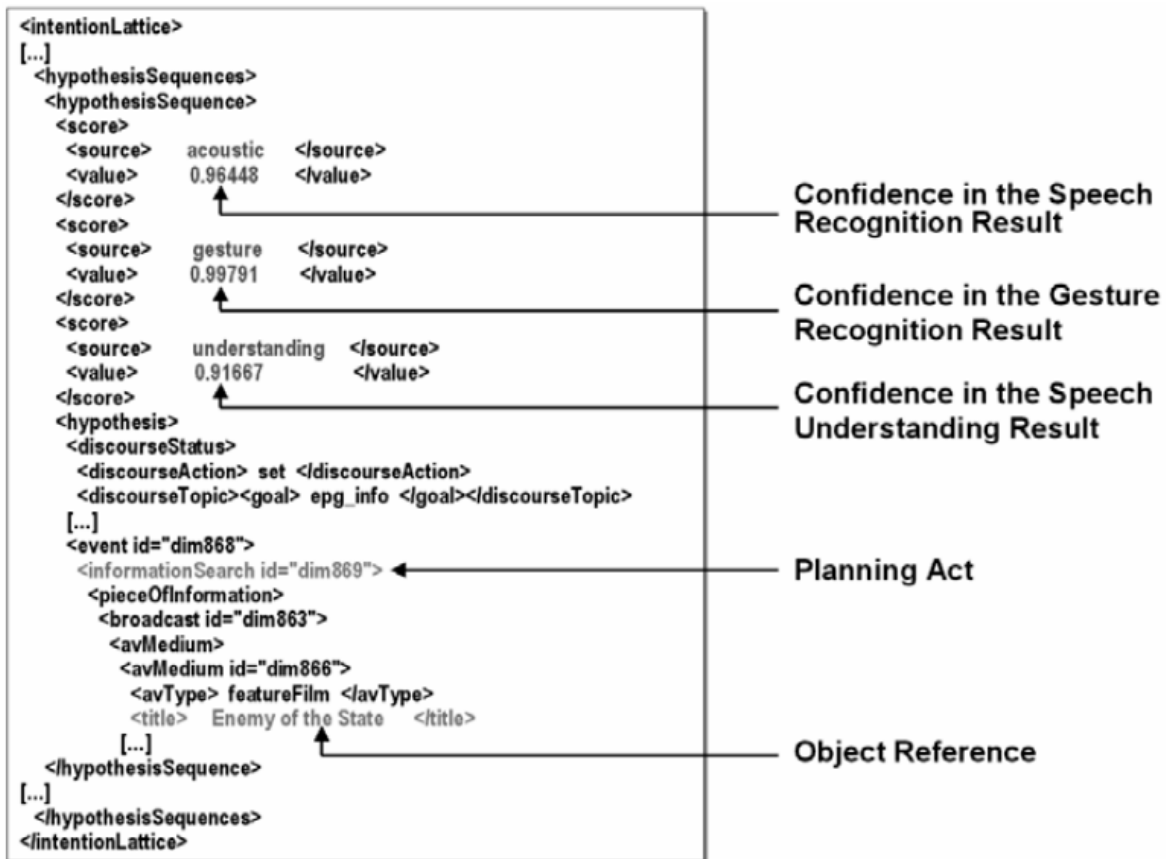
finally rank the remaining interpretation hypotheses and **to select the most likely one**, which is then passed on to the action planner.” (EX-1005, 14-15.)



SmartKom, Figure 2(60)

The “intention lattice” provided to the intention recognizer contains a number of hypotheses that “stand for alternative readings of one and the same user input.” (EX-1005, 287-288, 305; EX-1003, ¶¶120-122.) The intention lattice contains one or more intention segments that each includes, e.g., a discourse

object, goal manipulations (e.g., setting, retracting, or retaining a goal), slot manipulations (setting slots to establish new information, retracting to invalidate current information, or retaining slots to confirm information), annotation of the intention as positive or negative feedback (e.g., based on user state), and confidence scores from the analysis modules. (EX-1005, 306.) An exemplary intention lattice fragment is illustrated below. (EX-1005, Figure 8(15).)



SmartKom, Figure 8(15)

The information in the intention lattice for a hypothesis is continually updated as the lattice flows through the system. For example, the natural language

processing (speech interpretation) module “transform[s] the word lattice sent by the speech recognizer into a list of hypotheses representing ... possible user intentions.” (EX-1005, 195.) The hypotheses, each with corresponding recognizer and understanding score values, is provided to the modality fusion component which provides “integration and disambiguation of the different modalities.” (EX-1005, 204, 224.) Then, “a maximum of three hypotheses is sent to the intention analyzer”, each with a “fusion score.” (EX-1005, 232-233, 229.)

The intention recognizer engages a discourse modeler that works in conjunction with a context modeler to enrich and validate/score the received hypotheses. (EX-1005, 286-287, Figure 1(286) (below).) For example, the contextual domain (e.g., from the domain model), situational knowledge, and user profile/preference knowledge are used “for augmenting” the “intention hypotheses with implicit information” in the SmartKom-Kobsa system “to spell out their underlying intentions, and finally, to define a common background representation for the processed content, i.e., intention lattices.” (EX-1005, 271, 274.) A contextual coherence score for each hypothesis in the intention lattice is also provided to the intention recognizer. (EX-1005, 285.)

Using the scores and enriched hypotheses, the intention recognizer “select[s] the interpretation of the user input which is considered to be the best matching one.” (EX-1005, 287.) Specifically, the intention recognizer includes a

“probabilistic model” that “combines various scores, based on features in the representation and computed by the SmartKom modules” to support its selection of the intention hypothesis. (EX-1005, 285.) The selected hypothesis is an “*interpretation*” and therefore SmartKom determines “*an interpretation of the natural language utterance based on the identified context.*”

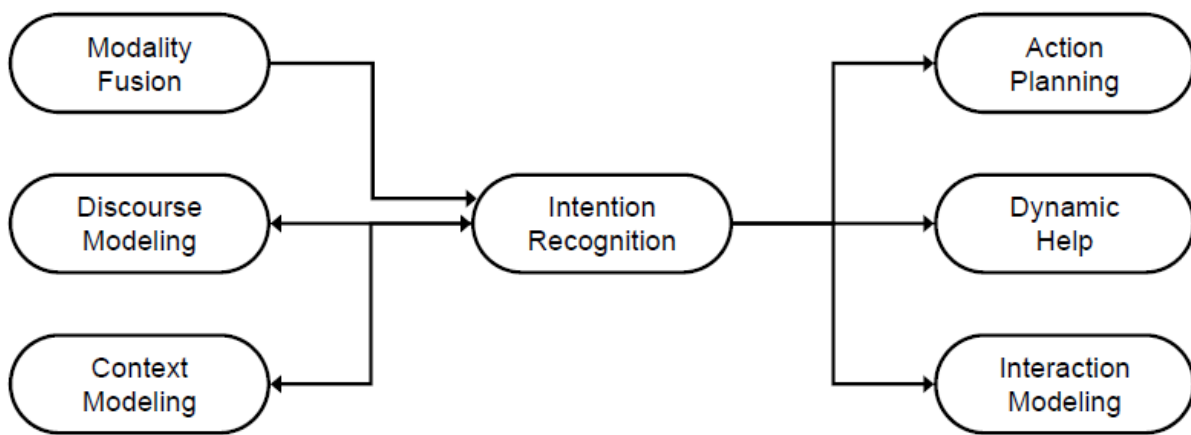


Fig. 1. Section of the multimodal interaction from the intention recognition point of view

SmartKom, Figure 1(286)

6. Short-Term and Long-Term Knowledge Limitations

It was well-known, prior to the '699 patent, that interactive dialogue systems used “information about the dialogue history and the user” “to constrain how the system interprets the user’s subsequent utterances and to determine what the system should say and how it should be said.” (EX-1020, 124.) These processes involve “representations of discourse structure, of intentions, goals and beliefs, and of dialogue as a collaborative activity.” (*Id.*) SmartKom-Kobsa similarly integrates

“*short-term*” and “*long-term*” knowledge to enrich hypotheses associated with potential interpretations of an utterance (“intention hypotheses”) with contextual information and to provide a contextual coherence score for each. (*See, e.g.*, §IV.B.5.) This knowledge is further used to adapt the system’s response. (EX-1005, 61, 323; §§IV.B.1, IV.B.7-8.)

Specifically, SmartKom’s architecture provides “shared knowledge services” including modeling components that “enable[] the system to act analogously, i.e., to provide hitherto implicit knowledge concerning what is talked about.” (*See, e.g.*, EX-1005, 277-278, Figure 2(60).) The specific knowledge stores and their associated contexts are illustrated in SmartKom’s Table 1(274) (below), that provides a “broad categorization of the types of context relevant to spoken dialogue systems, their content and respective knowledge stores.” (EX-1005, 274; EX-1014, 273-274.)

Table 1. Contexts, content and knowledge sources

Types of context	Content	Knowledge store
Dialogical context	What has been said by whom	Dialogue model
Ontological Context	World/conceptual knowledge	Domain model
Situational context	Time, place, etc.	Situation model
Interlocutionary context	Properties of the interlocutors	User model

SmartKom, Table 1(274)

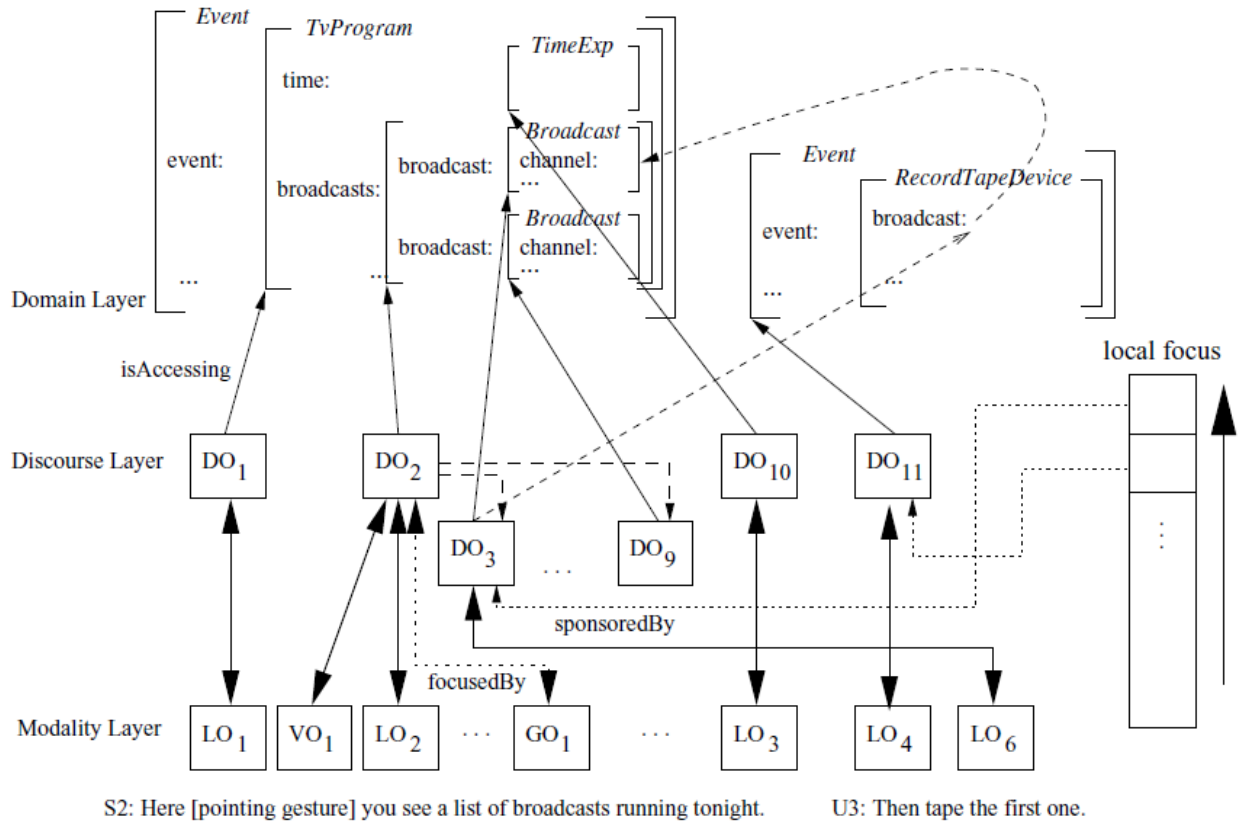
a) Accumulating Short-Term Knowledge [1E.1]-[1E.2]/ [12F.1]-[12F.2]

The SmartKom-Kobsa combination discloses a “*computer system*”/“*physical processor[s]*” that accumulates “*short-term knowledge based on one or more natural language utterances*” [1E.1]/[12F.1]. (EX-1003, ¶¶148-161.)

SmartKom’s dialogue model contains “what has been said by whom” and is associated with dialogical (discussion/conversation) context. (See EX-1005, Table 1(274).) That is, the dialogue model includes the dialogue history which is a type of “*short-term knowledge based on*” received “*natural language utterances,*” consistent with the ’699 patent’s disclosure. (EX-1003, ¶¶153-157; EX-1001, 5:10-12 (short-term knowledge includes “input received during a single conversation”).) In SmartKom, the dialog history is referred to as the “discourse state.” (EX-1005, 239, 242, 246.)

SmartKom accumulates dialog history (discourse state) and uses this knowledge to enrich and score intention hypotheses (representations of the

possible interpretations of an utterance). (See EX-1005, 237; EX-1003, ¶¶153-157.) In SmartKom, the “context representation” used “for representing the discourse state” consists of three levels: the discourse object layer, the modality layer and the domain object layer.” (EX-1005, 239, 242, Figure 2(243)(below).) The “discourse object layer is the central layer of the discourse representation” and “comprises the concepts introduced into the discourse.” (EX-1005, 242.) The modality layer includes modality objects (MOs), including linguistic objects (LOs) linked to discourse objects. (EX-1005, 244-245.) The local focus structure provides “access to all discourse objects” during interpretation of later utterances. (EX-1005, 245.)



SmartKom, Figure 2(243)

The discourse model “is dynamically updated as system output progresses.” (EX-1005, 61.) That is, the “*short-term knowledge*” contained in the discourse state stored in discourse memory is continuously accumulated during a conversation. (EX-1003, ¶157.)

The situation model, which records “time, place” of an utterance is also a type of “*short-term knowledge*” based on the received utterance, consistent with the ’699 patent’s disclosure. (EX-1003, ¶¶148, 158; EX-1001, 4:55-60 (indicating short-term knowledge includes “location” data).)

The SmartKom system also accumulates user state knowledge during a conversation which is “an extension of the well-known term of emotion with some internal states of a human like, e.g., ‘hesitant’, that are important in the context of human computer interaction (HCI).” (EX-1005, 139 (emphasis in original).) User state is also a type of “*short-term knowledge*” based on received utterances. (EX-1003, ¶159.) SmartKom accumulates the situational and user-state knowledge during a conversation. (EX-1003, ¶¶159-162.) For example, the prosody module takes speech signals and word lattices (word hypothesis graphs) from the speech recognizer and generates a “user state lattice.” (EX-1005, 141.)

The SmartKom-Kobsa combination discloses the “*short-term knowledge*” accumulates “*during a predetermined time period*” [1E.2]/[1F.2], consistent with the ’699 patent. The ’699 patent uses the term “*predetermined*” in the context of time only once and unrelated to the duration of the time period when the “*utterances*” are received:

For example, a human is unlikely to recall a context of a conversation from two years ago, but because the context would be identifiable by a machine, session context is **expired after a predetermined amount of time** to reduce a likelihood of contextual confusion based on stale data. However, relevant information from an expired session context may nonetheless be added to user, historical, environmental, cognitive, or other long-term knowledge models.

(EX-1001, 14:34-42.)

SmartKom-Kobsa discloses or suggests accumulating short term knowledge during a conversation between the user and system. SmartKom's discourse state is "based on the three-tiered context representation presented in Luperfoy (1992)."

(EX-1005, 239.) Luperfoy notes an "agent participating in a dialogue experiences information decay[ing] over the course of the conversation" with "information from the linguistic, discourse, and belief system tiers decays at different rates and in response to different cognitive forces/limitations." (EX-1012, 24-25.)

Specifically, LOs "become old and vanish at an approximately linear rate as a function of time counted from the point of their introduction into the discourse history, i.e., as LOs get older, they fade from the discourse and can no longer serve as linguistic sponsors for anaphors." (EX-1012, 25.) Discourse objects (referred to by Luperfoy as "pegs") "decay as a function of attentional focus, so that as long as an individual or concept is being attended to in the dialogue, the discourse peg will

remain near the top of the focus stack and available as a potential discourse sponsor for upcoming dependent referring expressions.” (EX-1012, 25.)

SmartKom stores its discourse state in “discourse memory.” (EX-1005, 242.) SmartKom acknowledges this decay of information, explaining that “[f]or longer dialogues (more than half an hour of discourse), the discourse memory runs out of memory” and in such cases, “even for most humans,” it is “necessary to forget information.” (EX-1005, 252.) Kobsa also discloses “at the end of a dialog session” (e.g., the conversation), the system “records all information about the user inferred from his/her dialog behavior in a corresponding file.” (EX-1006, 10-11.)

Thus, SmartKom-Kobsa discloses that “*short-term knowledge*” is accumulated for a dialog between the user and system over the shorter of (1) the duration of the conversation/dialog or (2) the size of short term storage (e.g., 30 minutes)—“*a predetermined time period.*” (EX-1003, ¶¶162-168.)

b) Accumulating Long-Term Knowledge [1F]/[12G]

The SmartKom-Kobsa combination discloses a “*computer system*”/ “*physical processor(s)*” that accumulates “*long-term knowledge*” “*based on one or more natural language utterances received prior to the predetermined time period*” [1F]/[12G]. (EX-1003, ¶¶169-176 (providing context for “*long-term knowledge*”).)

SmartKom’s Table 1(274) includes a “user model” storing the “properties of the interlocutors” (i.e., user information) and providing “interlocutionary” (or “user”) context as a knowledge source/store. (EX-1005, 274.) As explained by Kobsa, “long-term” user models “describe relatively static characteristics of users” such as “his/her knowledge about a domain” and the system’s “assumptions about the user’s experience with respect to a domain of discourse.” (EX-1006, 5, 17, 29, 52, 199; *see also*, 6, 200-205, 411-413.) Because the long-term user model includes knowledge derived from a user’s prior conversations, the long-term user model includes “*knowledge based on one or more natural language utterances received prior to the predetermined time period*” (i.e., utterances from previous conversations with the user). (EX-1001, 5:4-14 (long-term shared knowledge includes “explicit and/or implicit user preferences”).)

Table 1. Contexts, content and knowledge sources

Types of context	Content	Knowledge store
Dialogical context	What has been said by whom	Dialogue model
Ontological Context	World/conceptual knowledge	Domain model
Situational context	Time, place, etc.	Situation model
Interlocutionary context	Properties of the interlocutors	User model

SmartKom, Table 1(274)

SmartKom’s user model provides “properties of the interlocutors” to support interlocutionary/user context. (EX-1005, 274.) The knowledge stored in the user

model can be acquired (accumulated) “either *explicitly* or *implicitly*.” (EX-1006, 416 (emphasis in original).) For example, explicit knowledge can come from the application interviewing the user for certain information. (*Id.*) “Implicit acquisition involves ‘eavesdropping’ on the user-system interaction in order to observe the user’s behavior and from it to infer facts that go into the model.” (*Id.*) The user model infers user information from conversations with the user and therefore the stored user model includes knowledge about past conversations with the user.

Both SmartKom and Kobsa describe techniques for “*accumulating long-term knowledge*.” SmartKom describes an interaction model that “computes information on the user” from the system’s interpretation of an utterance. (EX-1005, 287.) Kobsa teaches the “*user modeling component* is the part of a dialog system whose function is to incrementally construct a user model; to store, update, and delete entries; to maintain the consistency of the model; and to supply other components of the system with assumptions about the user.” (EX-1006, 6 (emphasis in original).) As an example, Kobsa provides an overview of the General User Modeling System (“GUMS”) which “is designed for building long term models of individual users.” (EX-1006, 417.) Kobsa confirms the user model is persistent (long-term), teaching that, “at the end of a dialog session”, the system “records all information about the user inferred from his/her dialog behavior in a corresponding file.” (EX-1006, 10-11.)

Kobsa further teaches “at the end of a dialog session,” the system “records all information about the user inferred from his/her dialog behavior in a corresponding file.” (EX-1006, 10-11.) That is, the system updates the long-term user model with information before the session ends and short-term session knowledge is deleted.

7. Identifying Manner Of Speaking [1G]/[12H]

SmartKom-Kobsa combination discloses a “*computer system*”/ “*physical processor(s)*” that identifies “*a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge*” [1G]/[12H]. (EX-1003, ¶¶177-186.)

SmartKom’s “interaction module” identifies an indication of emotion mapped into models, the “*manner in which the natural language utterance was spoken.*” The interaction module applies a three-stage approach. First, “the emotional state of the user is recognized from facial expression and prosody.” (EX-1005, 320.) Second, “indications of problematic situations and **the emotional state of the user are collected from several sources and collectively evaluated.**” (*Id.*) The interaction module also “**analyzes the dialogue in respect to the style of interaction and the task and paradigm knowledge of the user.**” (*Id.*) “The interpretation of emotions and user states, and the generation of reactions to these states build the third stage.” (*Id.*) As discussed in §IV.B.8, the system adapts the

presentation of the response based on the identified manner. For example, “[i]f the system knows positive or negative preferences, it first presents objects that contain a preferred feature” and those “that show a disliked feature will be shown last.”

(*Id.*)

The interaction module “collects and evaluates indications of emotions, problematic situations and other aspects of the interaction.” (EX-1005, 321 (emphasis in original).) The component “operates by analyzing a set of possible *indicators*” (listed in Table 1(322) that “have values between 0 and 1.” (EX-1005, 321 (emphasis in original).) As shown in the Table below, the indications are derived from data collected from the prosody, user model, discourse model, and domain model.

Table 1. List of indicators

Source	Description
Mimic recognizer	Mimically conveyed anger
Prosody recognizer	Prosodically conveyed anger
Mimic recognizer	Mimically conveyed joy
Prosody recognizer	Prosodically conveyed joy
Mimic recognizer	Mimically conveyed dilatoriness
Prosody recognizer	Prosodically conveyed dilatoriness
Speech recognition	Linguistically conveyed anger
Speech understanding	Ratio of unanalyzable words
Intention analysis	Overall score of the best hypothesis
Intention analysis	Difference in score between first and second best hypotheses
Intention analysis	Number of possible hypotheses (depth of lattice)
Speech recognition	Score of the speech recognizer
Gesture recognition	Score of the gesture analyzer
Speech understanding	Score of the language analyzer
Media integration	Score of multimodal integration
Discourse history	Score of the discourse module
Domain model	Score of the domain module
Intention analysis	Final score of the intention module
Intention analysis	Number of elements in the user input
Discourse history	Number of new (not previously mentioned) elements
Speech understanding	Number of elements addressed by speech
Gesture analysis	Number of elements addressed by gesture
Media integration	Number of elements addressed by speech and gesture
Intention analysis	Importance of speech recognition score for overall score
Intention analysis	Importance of gesture analysis score for overall score
Intention analysis	Importance of domain model score for overall score
Intention analysis	Importance of language understanding score for overall score
Intention analysis	Importance of discourse model score for overall score
Speech understanding	Relative number of sentence-like units in one turn
Speech understanding	Relative number of words in one turn
Speech understanding	Relative frequency of pronouns
Speech understanding	Relative frequency of verbs
Speech understanding	Relative frequency of adverbs
Speech understanding	Relative frequency of nouns
Speech understanding	Relative frequency of content words
Speech understanding, language generation	Relative frequency of content words appearing in the system output
Speech understanding, language generation	Relative frequency of content words not appearing in the system output

SmartKom, Table 1(322)

The indicator values are then mapped to models “by means of a matrix multiplication.” (EX-1005, 322.) The interaction module “delivers four sets of models”: problem, user knowledge, modality, and linguistic. (EX-1005, 322-323.)

Table 2. List of models

Set	Description
Problem	Likelihood of a problem
Problem	Likelihood of an analysis problem
Problem	Discourse progress rate
Problem	Likelihood of the user being angry
Problem	Likelihood of the user being happy
UserKnowledge	Estimation of user familiarity with task
UserKnowledge	Estimation of user familiarity with system
Modality	Ratio of spoken input content
Modality	Ratio of gestural input content
Modality	Ratio of multimodal input content
ModalityContrastive	Ratio of contrastive usage of multimodal input
ModalityRedundant	Ratio of redundant usage of multimodal input
Linguistic	Adaptivity of user’s lexical choices to former system output
Linguistic	Likelihood of long turns
Linguistic	Likelihood of long sentences
Linguistic	Ratio of pronoun usage
Linguistic	Ratio of verb usage
Linguistic	Ratio of adverb usage
Linguistic	Ratio of noun and verb usage

The UserKnowledge set of model values “reflects the assumed task and paradigm knowledge of the user.” (EX-1005, 323.) “The task knowledge [“[e]stimation of user familiarity with task”] describes the user’s knowledge of the current task (e.g., programming a VCR), while the paradigm knowledge [“[e]stimation of user familiarity with system”] indicates how well the user is accustomed to dealing with multimodal dialogue systems, and, especially, with SmartKom.” (*Id.*) This UserKnowledge is obtained from the user model, which is “*long-term knowledge.*” (EX-1003, ¶183; §IV.B.6.b.)

Another set of models (“Linguistic”) describes “the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete

sentences and average length of input.” (EX-1005, 323.) These models “help to adapt the language generation in order to reflect the user’s style.” (*Id.*) Another set of models (“Modality”) “compares the use of different modalities by the user (for instance, a preference for gestures or spoken input).” (*Id.*) Three models in the final set identify user state including “likelihood that the user is angry” or happy. (*Id.*) The modality and linguistic sets and user state models include situational knowledge, discourse state, and user state information which are each “*short-term knowledge.*” (EX-1003, ¶185; §IV.B.6.a.)

Thus, the indication of emotions/user states mapped into models by the SmartKom-Kobsa system is the “*manner in which the natural language utterance was spoken*” which is collected and evaluated using the prosody, user model, discourse mode, and domain model (“*identif[ied] ... based on the short-term knowledge and the long-term knowledge*”).

8. Generating A Response [1H]/[12I]

SmartKom discloses a “*computer system*”/“*physical processor[s]*” that generates “*a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.*” [1H]/[12I]. SmartKom’s action planner, presentation planner, natural language generation [NLG] component, and speech synthesis cooperate to provide a response multi-modally to the user. (*See* EX-1005, 396, 401.)

SmartKom’s “action planner” generates a response to the user’s utterance in the form of an action plan. After selection of an interpretation from the set of hypotheses, the intention recognizer provides the interpretation to the action planner which makes “decisions on the applications to contact and the next interaction with the user.” (EX-1005, 287.) The action planner, also referred to as the dialogue manager, “*generat[es the] response to the natural language utterance based on the interpretation.*” (See EX-1005, 302, Figure 1(303)(below).)

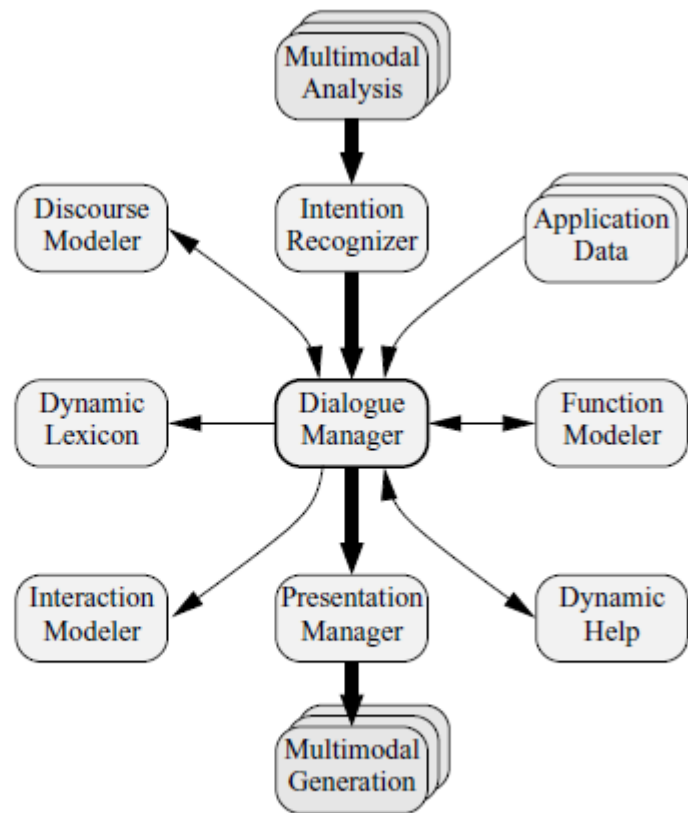


Fig. 1. Modules in direct communication with the dialogue manager

SmartKom, Figure 1(303)

The action planner “*generate[s] a response*” through development of an action plan “specify[ing] a possible course of subsequent communicative acts to reach” a goal articulated by the user in the utterance. (EX-1005, 311.) An action plan, illustrated below for responding the utterance “I want to send a document”, includes a set of steps (also referred to as “games”/“moves”). Each step includes the utterance/conversation type (e.g., instruct, inform, request-response, graphical action) and the channel which specifies the roles of the parties for the step. (*See* EX-1005, 305-313, Table 3(313)(below).)

Table 3. Fully expanded plan for sending a fax message

Step	Task	Channel	Application
1	Present clear screen	→ user (g)	fax
2-1	Present scanning area	→ user (g,s)	realDocument
2-2	Instruct to place the document	→ user (s)	
2-3	Initialize document scanner	→ realDocument	
2-4	Response initialization complete	← realDocument	
2-5	Request start scanning	→ realDocument	
2-6	Response scanning complete	← realDocument	
2-7	Request to remove document	→ user (s)	
2-8	Response document removed	← realDocument	
2-9	Request scanned image	→ realDocument	
2-10	Response scanned image	← realDocument	
2-11	Present scanned image	→ user (g)	
3-1	Present keypad and request number	→ user(g,s)	phone
3-2	Collect number	← user (reactive)	
3-3	Response number	← internal	
4	Inform “fax being sent”	→ user (s)	fax
5-1	Request transcribe picture	→ realDocument	realDocument
5-2	Response transcribed image	← realDocument	
6	Request sending of fax	→ telephony	fax
7	Response sending complete	← telephony	
8	Inform about completion of task	→ user (g,s)	

SmartKom, Table 3(313)

In the above plan, the “*interpretation*” is “I want to send a fax” which sets the goal of sending a fax. (See EX-1005, 312-313 (Table 3).) The responses to this utterance are based on the interpretation (“I want to send a fax”). Specifically, the system generates a response, e.g., to “[i]nstruct” the user “to place the document” (step 2-2) in the scanning area and later to request the user input a telephone number (step 3-1). (EX-1005, 313 (Table 3).)

The interaction module, described above in §IV.B.7, provides input to the presentation planner, natural language generation [NLG] component, and speech synthesis which cooperate to adapt the response provided by the action planner for presentation to the user. (See EX-1005, 396, 401.) Specifically, these components adapt the response based on the “*identified manner in which the natural language utterance was spoken.*”

As discussed in §IV.B.7, the interaction module, using sets of models, identifies the “*manner*” based on user knowledge, situational and discourse knowledge, and user state knowledge. SmartKom teaches the UserKnowledge models are employed by “the presentation/language generation to deliver an appropriate amount of feedback and assistance.” (EX-1005, 323.) The features identified by the linguistic model “help to adapt the language generation in order to reflect the user’s style.” (*Id.*) “The adaptivity of a user’s lexical choices to former system output is estimated, which can helping adapting dynamic language models used in SmartKom and language generation in order to maximize this value as a measure of the common vocabulary.” (*Id.*) The modality set of models allow “presentation and the behavior of the animated character” to be “adapted toward the user’s preferred distribution of the different modalities—users who prefer pointing gestures could be supplied with an interaction display with more possibilities and details, while users interacting mainly through speech should get

the most important system feedback also by spoken output.” (*Id.*) Finally, the system adapts the response based on user state information (e.g., whether user is angry or happy). (EX-1005, 324-325.)

Thus, the response provided is generated based on both “*the interpretation and the identified manner in which the natural language utterance was spoken.*”

C. Dependent Claims

1. Claims 2, 13

The SmartKom-Kobsa combination discloses “*the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.*” As discussed in §IV.B.7, the interaction module uses indicator values (e.g., indicators from the prosody module) to identify the “*manner.*” The “prosody module” identifies properties associated with an utterance, such as “**pitch, loudness, duration, speaking rate, and pauses.**” (EX-1005, 139.) From this data, “user state” is determined. (*See, e.g.,* EX-1005, 323, 346.)

The prosody module “has two main inputs: the speech signals from the *audio module* and the word lattices (*word hypothesis graphs*, WHGs) from the *speech recognizer.*” (EX-1005, 141 (emphasis in original).) The module includes an ANN classification block consisting of four independent classifiers used for the detection of “phrase Accents”, “phonetic phrase Boundaries”, “rising intonations

for Queries”, and “User states”. (EX-1005, 142.) The accents, phrase, rising intonations are each an “*indication of tone, pace, timing, [and/or] inflection.*”

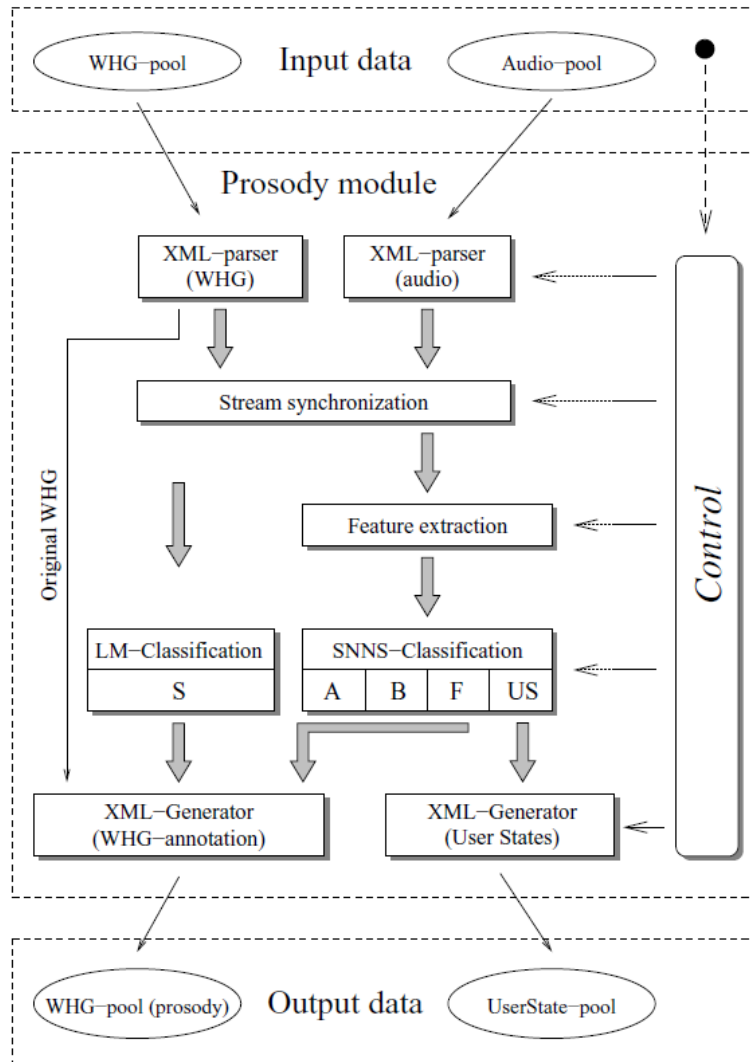


Fig. 1. Architecture of the prosody module

SmartKom, Figure 1(142)

As discussed in §IV.B.7, the SmartKom system includes an “interaction module” using, e.g., the linguistic model that “describes the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete

sentences and average length of input.” (EX-1005, 323.) These features “help to adapt the language generation in order to reflect the user’s style—based on the assumption that this is beneficial.” (*Id.*) “The adaptivity of a user’s lexical choices to former system output is estimated, which can helping adapting dynamic language models used in SMARTKOM and language generation in order to maximize this value as a measure of the common vocabulary.” (*Id.*) The linguistic model is an indication of “*word use.*”

2. Claims 3, 9, 14, 20

The SmartKom-Kobsa discloses “*the response comprises a voice response*” [3]/[9]/[14]/[20] and the response generation comprises “*vary[ing] one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response*” [3]/[14] and “*vary[ing] one or more of word use and/or jargon of the voice response*” [9]/[20].

SmartKom generates “*a voice response.*” In SmartKom, the “natural language generation (NLG) module” generates “not just plain text but rather generates complete syntactic structure and discourse information to supply highly annotated information structures for the concept-to-speech approach of SmartKom” for speech synthesis. (EX-1005, 401.) The speech synthesis module “not only produces audio data but also produces a detailed representation of the phonemes and their exact time points (in milliseconds) inside the audio signal.”

(EX-1005, 396.) This approach allows the character animation module to “generate[] a lip animation script for Smartakus that is then executed during speech output.” (*Id.*)

SmartKom’s “[i]nteraction modeling,” described in §§IV.B.7-8, “is concerned with different aspects like available modalities and user preferences for specific forms of communication as well as the affective state of the user.” (EX-1005, 61.) SmartKom’s “interaction model allows one to dynamically adapt the communicative behaviour of the system.” (*Id.*) For example, the model “describes the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete sentences and average length of input.” (EX-1005, 323.) All of these “help to adapt the **language generation** in order to reflect the user’s style.” (*Id.*) Kobsa also stresses that user models allow the system to “tailor[] object descriptions to the user’s level of expertise” and “adapt[] an expert system’s response behavior to the background knowledge of its users.” (EX-1006, 196.)

For example, SmartKom’s intention lattice “contain[s] additional information not directly relevant for dialogue planning, such as lexical items occurring in the input, which is passed on when triggering system output” and these items “enable[], e.g., the text generator component to **adapt to the user’s choice of words.**” (EX-1005, 306.)

Thus, the SmartKom-Kobsa combination varies “*word use*” and therefore discloses claims [3]/[9]/[14]/[20].

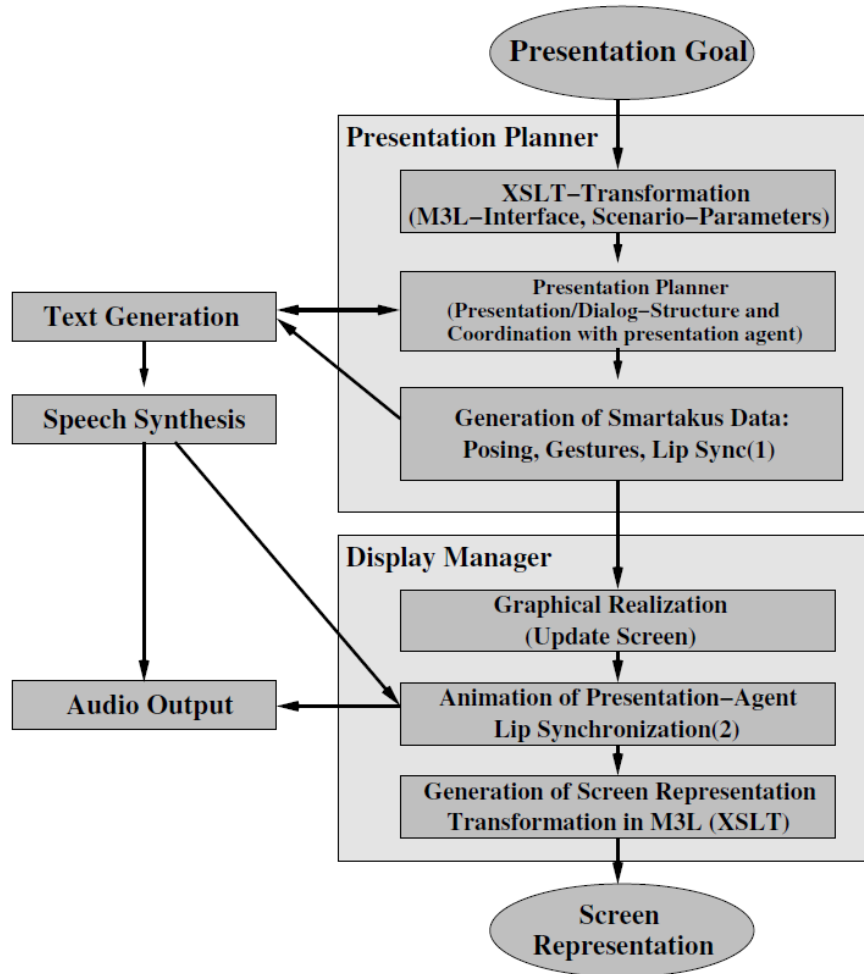
3. Claims 4, 15

SmartKom discloses a “*computer system*”/ “*physical processor[s]*” that obtains “*contextual signifiers and/or grammatical rules*” [4A]/[15A] and “*using the obtained context signifiers and/or grammatical rules [] generate sentences for use as response sets to cooperate with the user*” [4B]/[15B]. The ’699 patent does not provide an explicit definition of the term “*context signifier*.” However, based on specification, the plain meaning is any data (e.g., *short-term* and *long-term knowledge*) that is used to signify context of an utterance.

As discussed in §IV.B.6, SmartKom uses both short-term and long term knowledge to identify a context and this information is further used, as discussed in §IV.B.7, to identify a manner of speaking. Thus, SmartKom obtains “*contextual signifiers*.”

As discussed in §§IV.B.7-8, this information is used in the generation of a response (action plan and multi-modal output) to the user. More specifically, SmartKom uses gestures, mimics, speech, and graphics to present the response determined by the action plan to the user. (*See* EX-1005, 386.) “The presentation manager generates complete plans for graphics, gesture, and mimics output, while the plan for speech output is generated on an abstract high level only.” (EX-1005,

386, Figure 4(386)(below).) “The text generator creates a middle-level plan, i.e., a plan containing content-to-speech representations of the sentences to speak” and “the speech synthesis generates the low-level actions, i.e., the audio data (plus phoneme information that is used for lip synchronization. (EX-1005, 386-387.)



SmartKom, Figure 4(386)

The presentation planner “starts the planning process by applying a set of” presentation strategies. (EX-1005, 389.) In SmartKom, complex presentations depend on “input and modality parameters.” (EX-1005, 390.) For example, while

planning a response in the TV program context, “the presentation planner chooses the appropriate graphical element” based on amount and type of data to present and “take[s] into account what the user likes (e.g., if the user likes action films but not the channel ARD, action films will be shown first while broadcasts on channel ARD will be shown at the end or not shown at all).” (*Id.*) For the speech, “if the graphically presented information is [] the focus of a presentation, only a comment is generated for speech output” and “[i]f there is no graphically presentable information or it is insufficient to completely fulfill the presentation goal, more speech is generated.” (EX-1005, 387.)

Speech is generated by the natural language generation (NLG) module. (EX-1005, 401.) The system is “template-based” and also “based on lexicalized tree-adjointing grammar (LTAG) with full feature structures as its syntactic representation formalism.” (*Id.*) Specifically, the generation “is based on fully lexicalized generation” that “us[es] whole parts of a sentence together as one *fully specified template* that is then represented not as a strong but rather as a partial TAG derivation tree” illustrated below. (EX-1005, 402.) The set of LTAG trees is “considered a partial grammar.” (EX-1005, 404.) That is, the NLG module obtains “*grammatical rules.*”

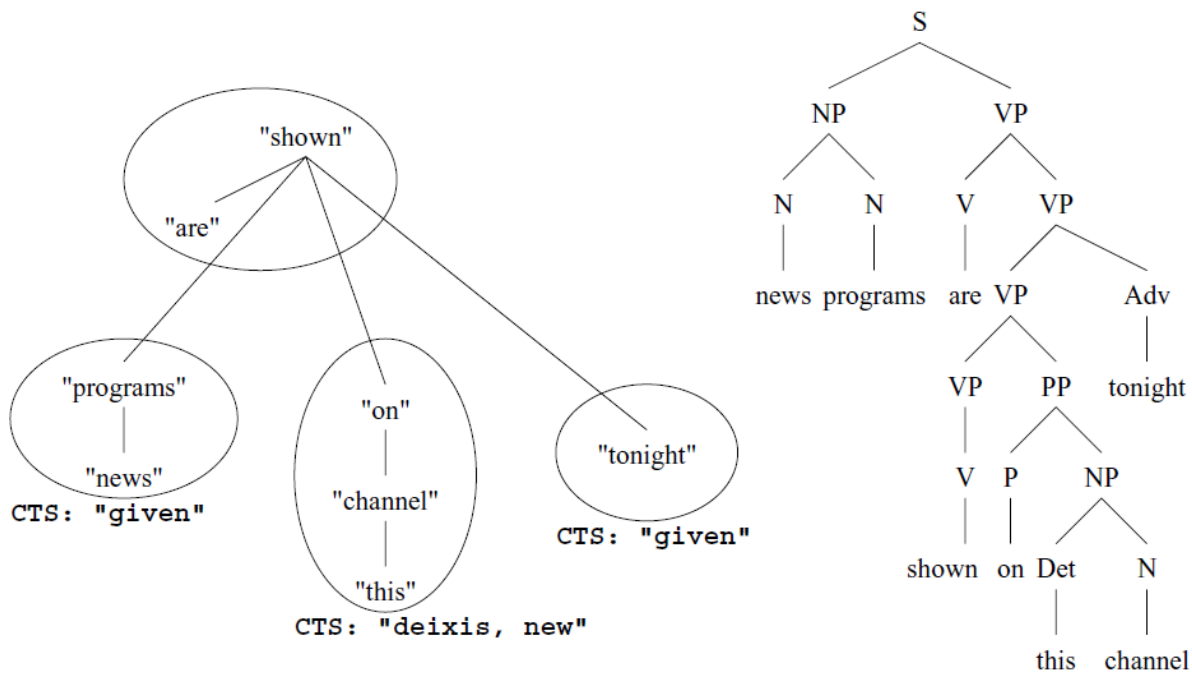


Fig. 1. Derivation tree with partial CTS (concept-to-speech) markup. Each ellipse is a fully specified template. The sentence-planning process combines such templates to a complete derivation tree, which when executed results in the derived tree shown to the *right* (shown without the feature structures attached to each node)

SmartKom, Figure 1(402)

The NLG generator, shown below, includes a microplanner that performs sentence planning by “map[ping] the domain-dependent relations of the abstract generator input onto more general semantic relations.” (EX-1005, 406.) “The semantic representation must also include additional information of pragmatic and/or situational content, such that this kind of information is available in further steps.” (*Id.*) “The representation at the semantics–syntax interface level is argument structure, i.e., a dependency tree enriched with feature annotations at the

nodes. (*Id.*) The “lexicalization” is performed at this stage. (*Id.*) In the syntactic realization phase, “the concrete syntactic representation for an utterance is generated” and the “[a]rgument structure is traversed and appropriate TAG elementary trees are found.” (*Id.*)

Thus, SmartKom “*us[es] the obtained context signifiers and/or grammatical rules [] generate sentences for use as response sets to cooperate with the user*” [4B]/[15B].

4. Claims 5, 16

The SmartKom-Kobsa combination discloses “*the long-term knowledge is associated with a first user*” and a “*computer system*”/“*physical processor[s]*” that generates “*a profile associated with the first user based on the long-term knowledge ...*” (*See* §IV.B.6.)

As explained by Kobsa, “[i]ndividual user models ... contain information specific to a single user.” (EX-1005, 414.) In the SmartKom-Kobsa combination, as described by Kobsa, the system “keeps individual models” and “thus will have a separate model for each user of the system.” (*Id.*) Thus, the “*long-term knowledge is associated with a first user.*” (*Id.*)

The SmartKom-Kobsa combination further generates “*a profile associated with the first user based on the long-term knowledge.*” Kobsa teaches the “*user modeling component* is the part of a dialog system whose function is to

incrementally construct a user model; to store, update, and delete entries; to maintain the consistency of the model; and to supply other components of the system with assumptions about the user.” (EX-1006, 6 (emphasis in original).) As an example, Kobsa provides an overview of the General User Modeling System (“GUMS”) which “is designed for building long term models of individual users.” (EX-1006, 417.) Kobsa also describes exemplary user profiles (Figure 3) accumulated through interpreting utterances in prior conversations reflecting “whether the user understands” basic concepts in the discourse domain. (EX-1006, 197, 15-17.)

Concept hierarchy of the system	User model	
	user's knowledge state	certainty of assumption
INFECTIOUS-PROCESS	KNOWN	100
HEAD-INFECTION	KNOWN	100
SUBDURAL-INFECTION	NOT-KNOWN	100
OTITIS-MEDIA	NO-INFORMATION	100
SINUSITIS	NO-INFORMATION	100
MENINGITIS	KNOWN	100
BACTERIAL-MENINGITIS	KNOWN	90
MYCOBACTERIUM-TB	NOT-KNOWN	70
VIRUS	NOT-KNOWN	90
FUNGAL-MENINGITIS	NOT-KNOWN	100
MYCOTIC-INFECTION	NOT-KNOWN	100
ENCEPHALITIS	NOT-KNOWN	90
SKIN-INFECTION	KNOWN	100

Figure 3. An example of an overlay model

Kobsa, Figure 3(16)

For the reasons discussed in §IV.B.4, the SmartKom-Kobsa combination also discloses “*the context for the natural language utterance is determined based further on the profile associated with the first user.*”

5. Claims 6, 17

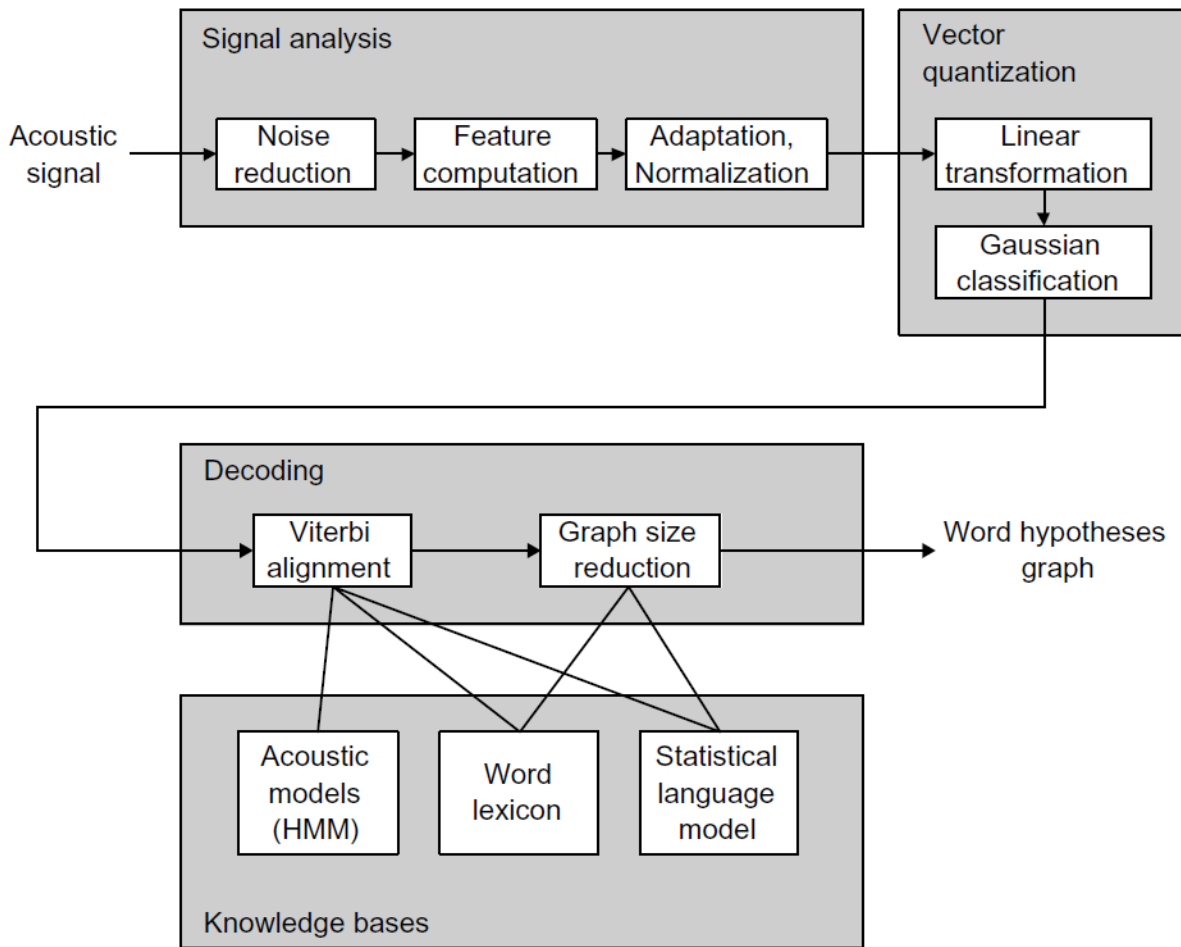
The SmartKom-Kobsa combination discloses a “*computer system*”/“*physical processor[s]*” that adapts “*the response based on a response format associated with the identified manner.*”

As discussed in §IV.B.8, the SmartKom system generates an action plan, including system responses, in a “*response*” plan. SmartKom’s “interaction model” working with the presentation planner, NLG component, and speech synthesizer “adapt[s] the communicative behaviour of the system.” (*Id.*) For example, the model “describes the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete sentences and average length of input” which “help[s] to adapt the **language generation** in order to reflect the user’s style.” (*Id.*) Kobsa also stresses that user models allow the system to “tailor[] object descriptions to the user’s level of expertise” and “adapt[] an expert system’s response behavior to the background knowledge of its users.” (EX-1006, 196.) Accordingly the “*response format*” is “*associated with the identified manner.*”

6. Claims 7, 18

SmartKom discloses a “*computer system*”/“*physical processor[s]*” that provides “*the natural language utterance as an input to a speech recognition engine*” and obtains “*the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.*”

As discussed in §IV.B.3, the SmartKom system includes a “**speech recognition engine**” that recognizes “*one or more words or phrases recognized from the natural language utterance.*” Specifically, the speech recognition engine receives “*as an input*” the “*natural language utterance*” in an “acoustic signal.” (EX-1005, 86, Figure 2(60), Figure 1(86)(below).) It then “transforms the acoustic signal into a **sequence of hypothesized words** in orthographic representation.” (*Id.*) The speech recognition engine adds a “confidence” score to each hypothesis and outputs this information to the speech interpretation (natural language understanding) component. (*See, e.g.*, EX-1005, 204.)



SmartKom, Figure 1(86)

7. Claims 8, 10, 19, 21

SmartKom discloses a “computer system”/ “physical processor[s]” that causes “the response to the natural language utterance to be provided to a user” [8]/[19] and adapts the response “to model a conversation” [10]/[21].

SmartKom “provides an anthropomorphic and affective user interface through an embodied conversational agent called Smartakus” which is designed to be polite and helpful, but never to appear to be obtrusive.” (EX-1005, 6, 304.) The

dialogue manager “plans and performs the dialogical interaction of Smartakus with the user.” (EX-1005, 301.) And “in determining [Smartakus’] actions, **realises the ‘personality’ of the system agent.**”(EX-1005, 304.)

SmartKom’s “interaction model” “adapt[s] the communicative behaviour of the system” (including Smartakus). (EX-1005, 61; §IV.B.6.) For example, the model “describes the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete sentences and average length of input” and “adapt[s] the **language generation** in order to reflect the user’s style.” (*Id.*) For example, system responses are further adapted based on user preferences, such “**the level of verbosity in speech output.**” (EX-1005, 407.) Kobsa also stresses that user models allow the system to “tailor[] object descriptions to the user’s level of expertise” and “adapt[] an expert system’s response behavior to the background knowledge of its users.” (EX-1006, 196.) That is, SmartKom is designed to adapt the response to “*model a conversation*” by “*adapting the response to have a personality by varying word use*” [10]/[21].

Specifically, as discussed in §IV.B.8, the action planner, also referred to as the dialogue manager, “*generat[es the] response.*” (See EX-1005, 302, Figure 1(303)(below).) The presentation planner, natural language generation [NLG] component, and speech synthesis cooperate to adapt and present the response

multi-modally to the user. (See EX-1005, 396, 401.) The example dialog discussed in §IV.B.4 illustrates the conversational response provided to the user:

U-1: *What's on TV tonight?*

SYS: (Displays the TV program listings) *Here you see a list of movies running tonight.*

U-2: *Is there a movie with Nicole Kidman?*

(EX-1005, 263.) Thus, SmartKom provides a response “*to a user*” [8]/[19].

8. Claims 11, 22

SmartKom discloses “*generat[e]/[ing] a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.*” As discussed in §IV.B.5, the SmartKom system generates a response based on the interpretation of the utterance.

As explained in §§IV.B.4, 8, SmartKom determines an interpretation of the utterances based on the identified “*context*” and generates a response “*sensitive to context.*”

In addition, as explained in §IV.B.6, SmartKom discloses utilizing “*user preferences*” based on “*short term knowledge and long term knowledge*” in identifying the user’s manner of speaking and generating the response.

As also explained in §IV.B.6, the SmartKom system accumulates short-term knowledge that includes a dialog history of “*words uttered by the user in one or*

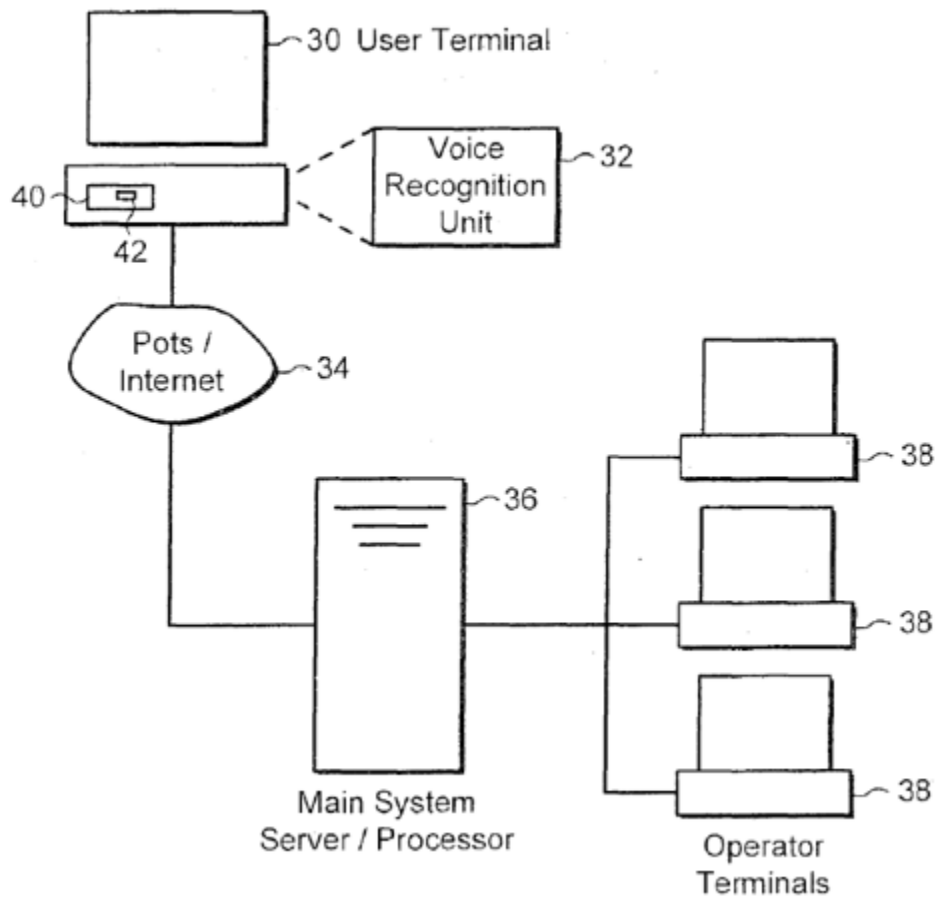
more prior natural language utterances.” This information is used, as discussed in §§IV.B.7-8, to generate the response to the user.

V. GROUND 2: Barbara-Ross-Kellner Combination Renders Obvious Claims 1-22

A. Combination Overview

1. Barbara

Barbara’s system (Figure 6, below) interprets an utterance “to determine what the user wanted, i.e. the intent of the spoken word.” (EX-1007, ¶82.) The system includes voice recognition unit 32 “operable to receive the spoken word and translate it into an electronic format” that is passed to processor(s) “having software for receiving, interpreting and correcting the incoming information.” (EX-1007, ¶¶ 83-85.)



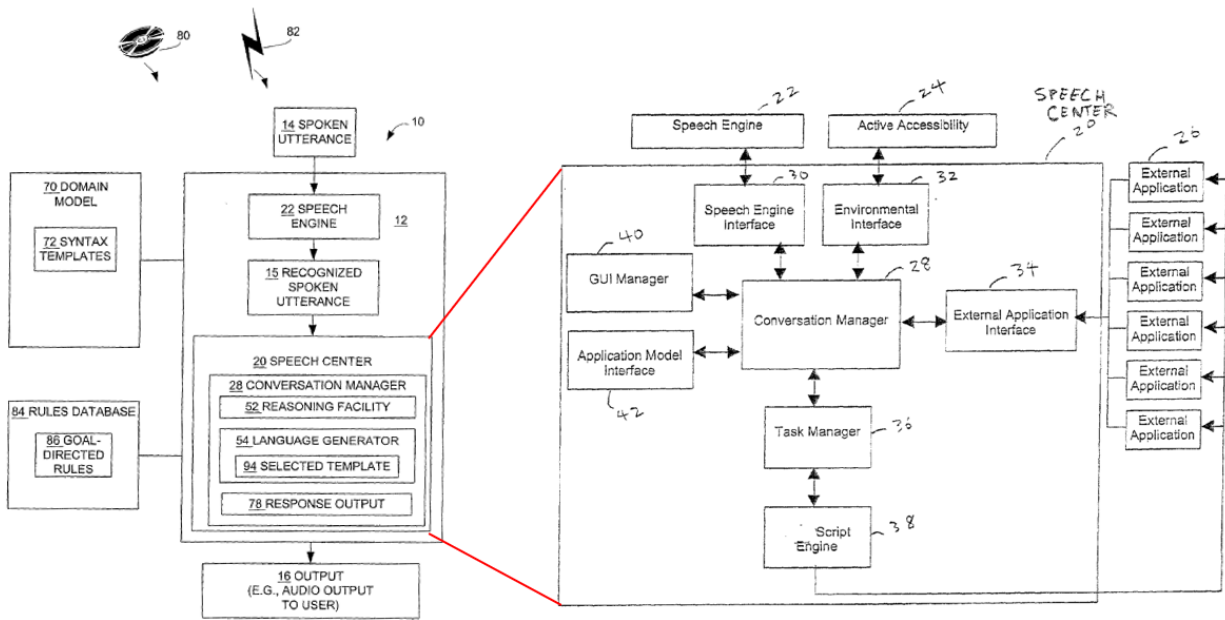
Barbara, Figure 6

When a user's "utterance has been detected," voice recognition unit 32 provides its "best guesses" "as to what was said," and what was meant. (EX-1007, ¶86.) After the correct intent is determined, the system responds. (EX-1007, ¶¶105-107, Figure 7.)

Utterances can be manually corrected if they "fail[] to meet the pre-determined criteria for automatic acceptance," (EX-1007, ¶¶100-104), and Barbara recognizes this "will be used heavily" during" an "initial period." (EX-1007, ¶102.)

2. Ross

Ross discloses a “conversation manager [that] processes spoken utterances” and “develops responses.” (EX-1008, Abstract.) System 10 comprises processor 12 including speech engine 22 and speech center 20 having conversational manager 28. (EX-1008, ¶¶22-24, Figures. 1-2 (below).)

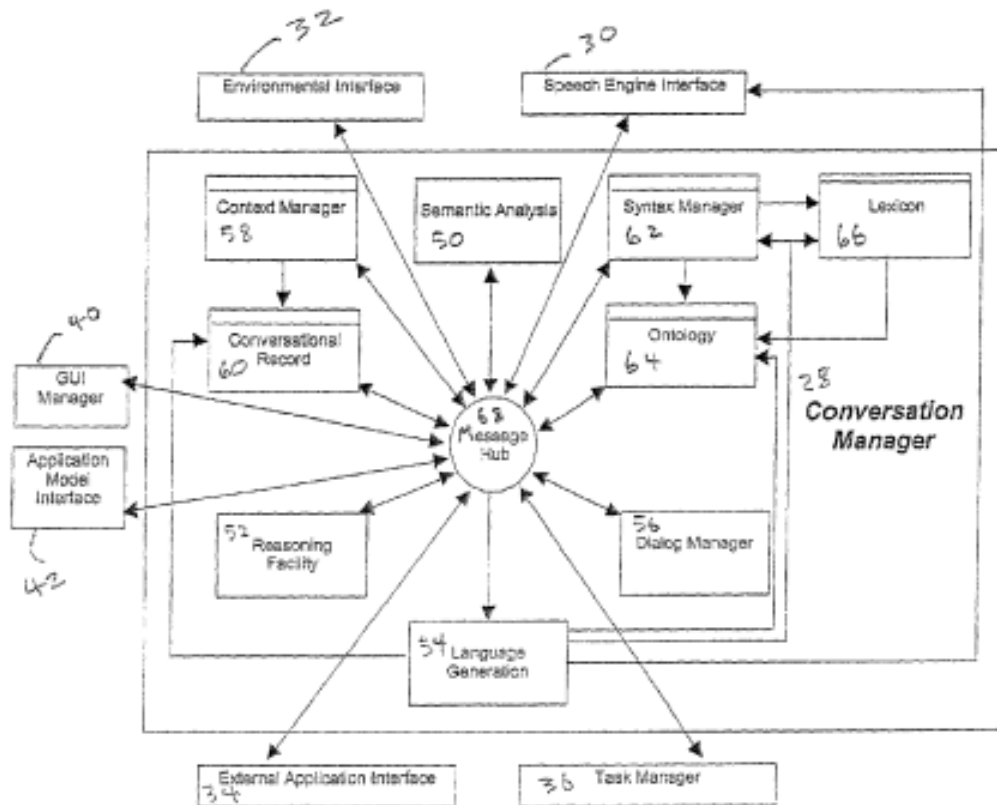


Ross, Figure 1

Ross, Figure 2

Conversational manager 28 includes reasoning facility 52 that “performs the reasoning process” and is concerned with “how to achieve the goals derived from the user’s questions and command.” (EX-1008, ¶56, Figure 3 (below).)

Conversational record 60 (part of the conversation manager) stores each utterance in a conversation and domain model 70. (EX-1008, ¶57.)



Ross, Figure 3

Because “[c]onversational speech is full of implicit and explicit references back to people and objects that were mentioned earlier,” (EX-1008, ¶57), Ross’s systems uses the conversational records and domain model to process utterances. (EX-1008, ¶¶22, 57.)

3. Kellner

Kellner discloses “a dialog system... comprising processing units for automatic speech recognition..., natural language understanding..., derived from user inputs, [and] generating acoustic and/or visual system outputs. (EX-1023, Abstract.)

In Kellner's dialog system, "the contents and form of the system outputs **are adapted to the style of speech of user inputs and/or to the behavior of a user.**"

(EX-1023, ¶5.) This is achieved using an "**associated user model.**"

Once "the style of speech of a user input is determined," various "characterizing features" can be evaluated (discussed below). (EX-1023, ¶¶15-22.)

"Also a user's interaction behavior is incorporated in the associated user model so that, more particularly, the output modalities used by the dialog system during a dialog are used." (EX-1023, ¶5.) Specifically, "[i]n dependence on the determined style data, the style of speech outputs is adapted, i.e. respective polite phrases and the same address are used, the speech level is adapted to the detected speech level; in dependence on the information density in a speech input more or fewer extensive system outputs are generated and the vocabulary used for the speech outputs is selected accordingly." (EX-1023, ¶23.)

As a result, "[t]he dialog system is thus in a position to generate system outputs adapted to a user's style of speech," namely "speech considered pleasant by the respective user." (EX-1023, ¶5.)

4. Motivation to Combine Barbara With Ross

While Barbara discloses a "server" to process utterances, it lacks details regarding the server's architecture. A POSITA would have been motivated to

combine Ross’s teachings regarding the architecture of a conversation manager with Barbara’s voice interface system. (EX-1003, ¶218.)

Barbara and Ross are analogous art to the ’699 Patent. Each pertains to the same field of endeavor, i.e., speech recognition systems. (*Compare* EX-1001, 71:28-29 *with* EX-1007, ¶7 *and* EX-1008, Abstract; EX-1003, ¶219.)

Barbara motivates the combination. For example, Barbara discloses the system “may either speak back or perform some action ... or both” in response to an utterance. (EX-1007, ¶107.) But, Barbara lacks details about how to generate such a response. A POSITA would have been motivated to search for a reference with such implementation details and would have been led to Ross which provides robust disclosure of its conversation manager including response generation. (EX-1003, ¶220.) That is, a POSITA would have been motivated to combine Barbara with Ross to provide a system with flexible response generation capabilities.

While Barbara discloses use of an “information database” and “knowledge base” to interpret utterances, Barbara provides limited detail regarding managing the information in those data stores. Ross discloses creating conversational records for storing a dialog history between the user and system including each user utterance and the system’s interpretation. (EX-1008, ¶57.) Ross further teaches that this conversational information “is eventually purged from the conversational record when it is no longer relevant to active goals.” (*Id.*) Implementing Ross’s

teachings regarding purging information from the “conversational record” allows Barbara’s dialog history to be expired and long-term data to be retained. (EX-1003, ¶221.) A POSITA would understand this data management approach improves storage efficiency and reduces need for extensive hardware storage. (EX-1003, ¶221)

Finally, the combination is nothing more than the application of a known technique (Ross’s conversation manager and data storage) to a known device (Barbara’s voice interface system) ready for improvement for the reasons discussed above. (EX-1003, ¶222.)

A POSITA would have had a reasonable expectation of success and the results of the combination would have been predictable. The Barbara-Ross combination would merely implement Barbara’s knowledge base to use Ross’s conversation records which expire, releasing memory space and Ross’s conversation manager teachings including response generation components. (EX-1003, ¶223.) Integrating Ross’s teachings into Barbara’s system would have been well within the skill set of a POSITA because the combination involves software and data storage, concepts well understood before the earliest priority date of the ’699 patent. (EX-1003, ¶223.)

5. Motivation To Combine Barbara And Ross With Kellner

A POSITA would have been motivated to combine Kellner's teachings of "adapt[ing] to the style of speech of user inputs and/or to the behavior of a user during a dialog with the dialog system" (EX-1023, ¶5) with Barbara's "voice interface where the end system responds to spoken natural language commands" (EX-1007, ¶82), as modified by Ross to enable Barbara's system to be "in a position to generate system outputs adapted to a user's style of speech with a style of speech considered pleasant by the respective user" (EX-1023, ¶5).

A POSITA would have been motivated to make this combination for numerous reasons. First, Kellner explicitly motivates the combination, because it would create a style of speech that is pleasant to the user. Kellner discloses that based "on the determined style data, **the style of speech outputs is adapted.**" (EX-1023, ¶23.) As a result, "[t]he dialog system [can] **generate system outputs adapted to a user's style of speech with a style of speech considered pleasant by the respective user.**" (EX-1023, ¶5.)

Second, a POSITA would have adapted the style of speech output to a particular user's speech style as taught by Kellner, because that would create a pleasant experience by providing results tailored to a user. "The dialog system [can] generate system outputs adapted to a user's style of speech with a style of speech considered pleasant by the respective user. **As a result, the inhibition**

threshold of the use of the dialog system can be lowered.” (EX-1023, ¶5.) A POSITA would have understood that any threshold that would inhibit, e.g., prevent, the use of the system, would be lowered. (EX-1003, ¶226.) Therefore, the system would capture additional user information as the user interacts more with the system, and therefore, provide results tailored to the user’s preferences and needs.

Third, the combination is nothing more than use of a known technique (adapting an output of a dialog system to the style of speech of user inputs and/or to the behavior of a user during a dialog with the dialog system) to improve a similar device (Barbara’s voice interface system) in the same way (by adapting the output of Barbara’s voice interface system to the user’s style of speech and/or behavior).

Third, Barbara, Ross, and Kellner are analogous art to the ’699 Patent. All three pertain to the same field of endeavor, i.e., speech recognition systems. *Compare* EX-1001, 1:28-29 *with* EX-1007, ¶¶7, 12, 82 *and* EX-1008, ¶Abstract *and* EX-1023, ¶1.

A POSITA would have had a reasonable expectation of success in the combination and the results of the combination would have been predictable. Combining Barbara and Ross with Kellner would merely implement Kellner’s “user model” into Barbara’s information database to capture the user’s style of

speech, such that the output of Barbara's system can be adapted according to the user's style of speech. (EX-1023, ¶¶5, 15-22). Such modification would merely reflect how to implement Barbara's response system and information database(s) and would have been well within the skill set of a POSITA before the earliest priority date of the '699 patent. (EX-1003, ¶¶224-229.)

B. Independent Claims

1. Preambles [1P.1]/[1P.2]/[12P]/[12A]

The Barbara-Ross-Kellner combination discloses [1P.1]/[1P.2]/[12P]/[12A]. Barbara discloses a “*method*” [1P] and “*system*” [12P]: “**system and method** for improving accuracy of signal interpretation.” (EX-1003, ¶¶231-234; EX-1007, ¶¶5, 7, 33-82, 86, Figures 1-7.)

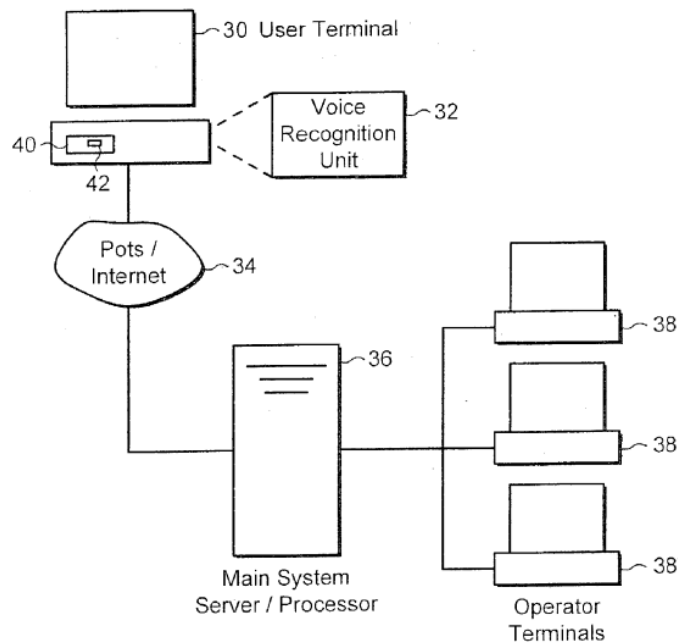
Barbara's system “*generat[es] natural language system responses.*” (EX-1003, ¶232.) For example, Barbara's system is “used to allow text that is spoken into voice recognition equipment to create a voice interface where the end system **responds to spoken natural language commands.**” (EX-1007, ¶82.)

Barbara's system has a main server with “**one or more service processors**” [*one or more physical processors*] including “**software** for receiving, interpreting and correcting the incoming information from the user terminal.” (EX-1007, ¶85, Figure 6 (below).) Although Figure 6 shows voice interface equipment 32 at a user terminal, in an embodiment, end-user's terminal 30 is “used as a telephone” with

voice recognition software “provided at a remote location.” (EX-1007, ¶83.) Thus, the software associated with voice recognition is also included in the main server, and Barbara’s “method” is “computer-implemented.” (EX-1003, ¶¶231, 234.)

Barbara therefore discloses “**a computer program** [“computer program instructions”]... for controlling [the] method of interpreting electronic signals, the computer program comprising **instructions.**” (EX-1007, ¶24, claim 30.)

[1P.2]/[12A].



Barbara, Figure 6

Kellner discloses “the contents and form of the system outputs **are adapted to the style of speech of user inputs** and/or to the behavior of a user during a dialog with the dialog system.” (EX-1023, ¶5.) Specifically, Kellner discloses “[i]n

dependence on the determined style data, **the style of speech outputs is adapted.**”
(EX-1023, ¶23.) [1P.1]/[12P.1].

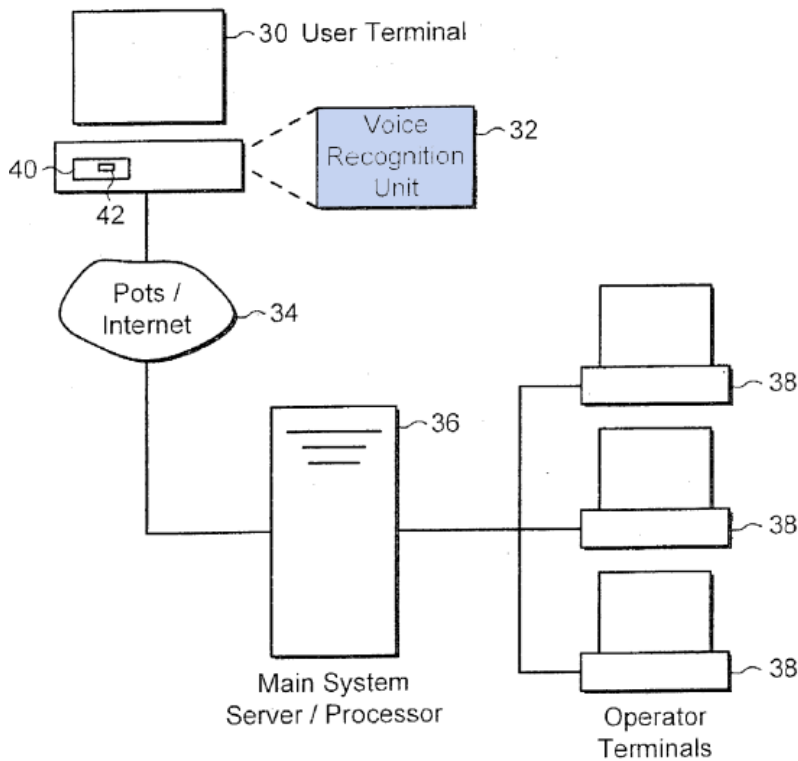
2. Receiving An Utterance [1A]/[12B]

Barbara discloses limitations [1A]/[12B].

Barbara’s system includes voice interface equipment 32 (shaded blue)

“**operable to receive the spoken word** and translate it into an electronic format.”

(EX-1007, ¶¶83, 86, Figure 6 (below).) The voice recognition software, whether in the user device or in the server, alone or in combination with the user device (e.g., a microphone) is an input device that “*receive[s] a user input that comprises a first natural utterance.*”

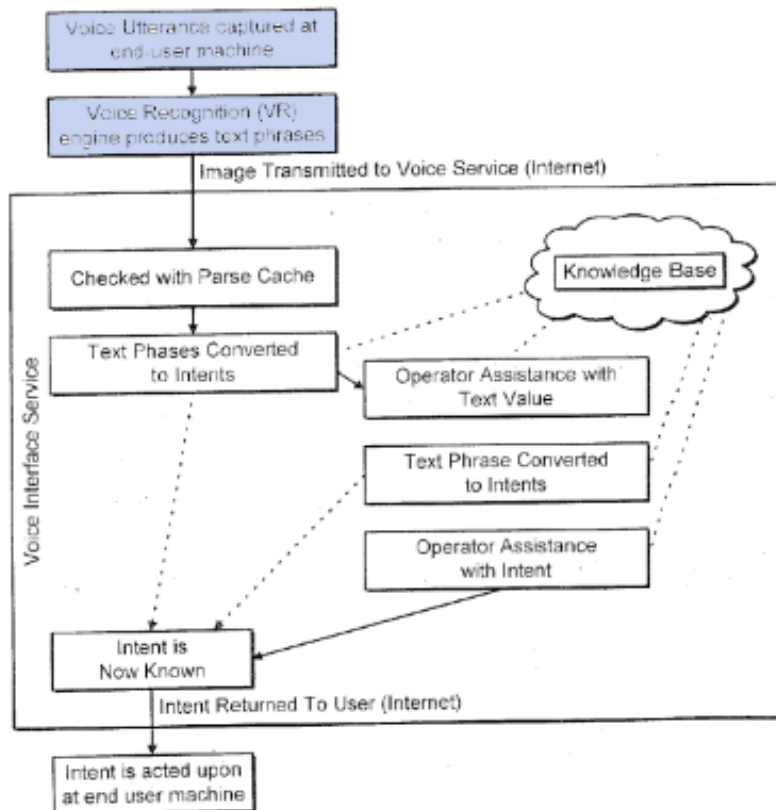


Barbara, Figure 6

3. Recognizing Words or Phrases [1B]/[12C]

Barbara discloses limitations [1B]/[12C].

Barbara's system is used to "create a voice interface where the end system responds to **spoken natural language commands.**" (EX-1007, ¶82.) Specifically, Figure 7 (below) shows that "when the end-user says something at their machine, the "end-user's machine and the voice recognition/interface service [] collectively recognize that **the user has said something** and determine[s] what has been said." (EX-1007, ¶86.) For example, "the software asks the voice recognition engine **for its recognized text phrases.**" (EX-1007, ¶86.)



Barbara, Figure 7

4. Identifying a Context [1C]/[12D]

Barbara discloses limitations [1C]/[12D].

Barbara explains “[c]ontextual information ... is captured as the user interacts with the system, in much the same way as interactions work in real life.” (EX-1007, ¶98.) For example, if “the user asks for the date of Toto’s last flea vaccination, it is much more likely that they are referring to Toto the dog, rather than Toto the restaurant.” (EX-1007, ¶148.) That is, Barbara teaches context (pets versus restaurant) is identified from recent dialog history which includes a reference associated with the context-domain pets (e.g., “flea vaccination”) (“*one or more words or phrases*”). (EX-1003, ¶¶235-238.)

5. Determining an Interpretation [1D]/[12E]

The Barbara-Ross-Kellner combination discloses limitations [1D]/[12E].

Barbara’s system interprets the utterance “to determine what the user wanted, i.e. the intent of the spoken word.” (EX-1007, ¶82.) Specifically, the utterance is “evaluated into a set of possible requests that the user may have actually meant.” (EX-1007, ¶¶90-93) “Contextual information” is then used to “evaluate[] the likelihood that the user really is asking each of those things.” (EX-1007, ¶98.) That is, Barbara selects one of multiple potential meanings based on the context and in doing so, “[a]mbiguities are resolved.” (*Id.*)

For example, “[i]f the user asks for the date of Toto’s last flea vaccination,” Barbara’s system uses the identified context domain (“pet”) to determine “it is much more likely that they are referring to Toto the dog, rather than Toto the restaurant.” (EX-1007, ¶148.) Accordingly, “*an interpretation of the natural language utterance*”, e.g., the word Toto refers to a dog, is “*determined ... based on the identified context*,” e.g., “pet.”

6. Short-Term Knowledge [1E.1]/[1E.2]/[12F.1]/[12F.2]

In the Barbara-Ross-Kellner combination, the “best guesses of the voice recognition engine as to what was said are evaluated” in the conversational speech processing component “against an information database(s) and knowledge base.” (EX-1007, ¶88.) The information database stores “personal details or preferences from what the user has said” and includes non-personal information, such as “weather reports or stock prices.” (EX-1007, ¶89.) Barbara also discloses “[t]he correct utterance text and the correct intention are stored in the knowledge base.” (EX-1007, ¶106.) Both are included in the main server. (EX-1007, ¶89, Figure 7.) Barbara’s system uses the information from both the information database and knowledge base to evaluate possible requests that the user may have meant. (EX-1007, ¶¶90-93.)

The Barbara-Ross-Kellner combination uses Ross’s data organization, for example the knowledge base is short-term storage that uses Ross’s conversational

records to store (“*accumulat[es]*”) each utterance “along with the results of its semantic analysis.” (EX-1008, ¶57; EX-1007, ¶106.) This information “is eventually purged from the conversational record when it is no longer relevant to active goals and after some predefined period of time has elapsed.” (EX-1008, ¶57.) That is, the “*short-term knowledge*” is accumulated in the conversational record during the duration of the conversation, e.g., when relevant to active goals, or for a predefined period of time—“*a predetermined time period.*” Thus, the Barbara-Ross-Kellner combination discloses “*short-term knowledge based on one or more natural language utterances received during a predetermined time period*” [1E.1]/[12F.1]. (EX-1003, ¶¶239-241.)

Barbara discloses, in a medical context, that “[a] typical expert system works by asking the user questions, and continuing to ask relevant questions, narrowing down the diagnosis until a conclusion of satisfactory confidence is reached.” (EX-1007, ¶168.) A POSITA would have understood that during the course of the diagnosis the one or more natural language utterances would be related to a single conversation because the user’s interaction with the computer system had one goal, e.g., to narrow down the diagnosis. Accordingly, Barbara alone discloses *the one or more natural language utterances are related to a single conversation between a user and the computer system.* (EX-1003, ¶243.)

Additionally, in the Barbara-Ross-Kellner system, during a *predetermined time period*, “each utterance is indexed in the conversational record 60, along with the results of its semantic analysis” until “it is no longer relevant to active goals.” (EX-1008, ¶57.) Accordingly, a POSITA would have understood that the received utterances that are indexed within the conversational record, would be related to a single conversation, e.g., a conversation associated with the user’s current “active goals.” (EX-1003, ¶¶244-245.)

7. Long-Term Knowledge [1F]/[12G]

Barbara records (“*accumulates*”) “**personal details or preferences from what has been said**” in the “information database(s).” (EX-1007, ¶88; EX-1007, ¶98 (“context information ... is captured as the user interacts with the system”.) For example, Barbara’s system infers that a speaker “who asks about their dog ten times a day” is a dog owner (personal detail) and therefore “might be considered very likely to be referring to it” during a future conversation. (*See, e.g.*, EX-1007, ¶98.) As a further example, based on stored user profile data, the system determines that “[s]omeone who is not involved in financial markets is unlikely to be asking about the Nikeii”, implying the system infers preference/employment information about the user. (EX-1007, ¶98.) This information, derived “from what has been said,” is knowledge about one or more past conversations with the user, and therefore is “*knowledge based on one or more natural language utterances*

received prior to the predetermined time period.” (EX-1003, ¶¶246-249.) Because this content “is built up over time”, the accumulated information is “*long-term knowledge.*” (EX-1003, ¶247; EX-1001, 5:10-13 (“Long-term shared knowledge may **include explicit and/or implicit user preferences**”).)

The Barbara-Ross combination discloses limitations [1F]/[12G].

8. Identifying a Manner [1G]/[12H]

The Barbara-Ross-Kellner combination discloses limitations [1G]/[12H].

Kellner identifies numerous manners in which an utterance can be spoken, including:

- number of polite phrases used,
- address used (you),
- speech level (colloquial language, standard language, dialect)
- information density (number of words recognized as significant of a speech input in relation to the total number of words used),
- ...number of different words in user inputs,
- classification of words of speech inputs with respect to rare occurrence.

(EX-1023, ¶¶15-22.)

In Barbara-Ross-Kellner, the “style of speech of a user input” [*a manner in which the natural language utterance was spoken*], is therefore evaluated based on different characterizing features, which include, among others, the “speech level” of the user input, e.g., one or more user’s utterances.

Kellner also discloses “[t]he details about the style of speech and dialog interactions of a user are contained in an **associated user model** which is evaluated by the dialog system components.” (EX-1023, ¶5.) Specifically, “[w]hen user models are generated, more particularly the style of speech and interactions occurring between user and dialog system are taken into account.” (EX-1023, ¶14.) As explained in §V.A.4, Ross discloses a conversational record for storing current-conversation utterances and their interpretation. In Barbara-Ross-Kellner, a POSITA would have found obvious to store “[t]he details about the style of **speech** and dialog interactions” (EX-1023, ¶5) of the current conversation in Ross’s conversational record—and then expire the information into the Kellner’s user model—because as explained in §V.A.4, this data management approach improves storage efficiency and reduces need for extensive hardware storage.

A POSITA would have understood “*identifying [by the computer system] a manner in which the natural language utterance was spoken*” is “*based on the short-term knowledge*” (characterizing features of the current user input/utterance) “*and the long-term knowledge*” (user model) because, for example, in the case of the identifying a user “speech level,” Barbara-Ross-Kellner’s system would compare the speech level of the user input (current utterance/conversation) with the speech level stored within the associated user mode, so that the system can compare, for example, if the user is using colloquial language versus standard

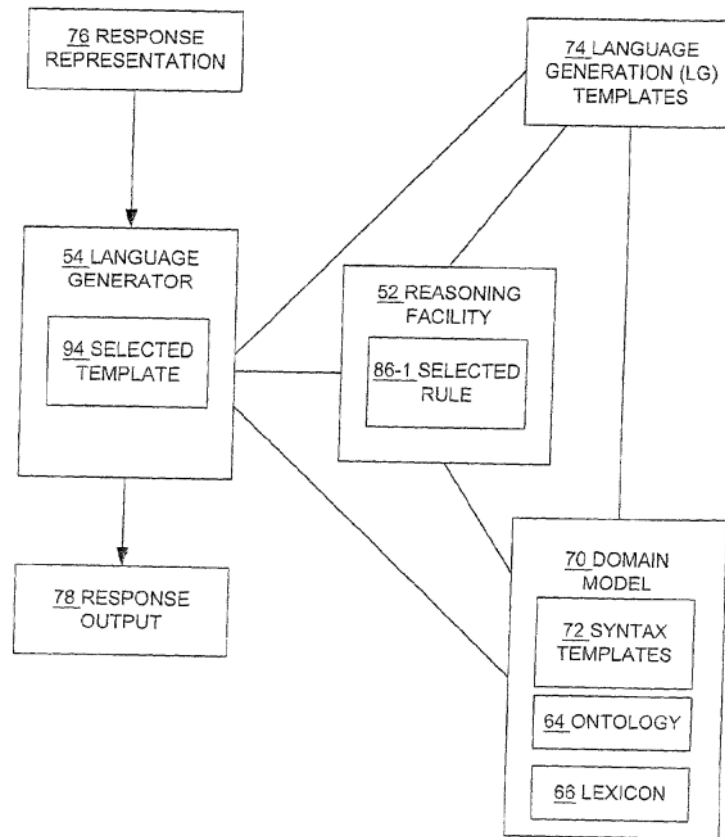
language. That is, the user model includes information about the speech levels (e.g., colloquial language, standard language, dialect) of its associated user, so the system can compare whether the speech level of a current utterance maps closer to any of the colloquial, standard, or dialect speech levels. (EX-1003, ¶¶250-253.)

9. Generating A Response [1H]/[12I]

The Barbara-Ross-Kellner combination discloses limitations [1H]/[12I].

In Barbara, after the intended meaning is established, the system “performs the correct intention,” e.g., by generating a response to the user. (EX-1007, ¶107.) For example, in Barbara, “feedback is given by the application visibly responding” or “the appliance speak[ing] back or perform[ing] some action or both.” (EX-1007, ¶107; *see also* EX-1007, ¶¶119-62.)

In Ross, the reasoning facility within the conversation manager generates a response representation 76 (set of propositions) “in response to the recognized spoken utterance 15.” (EX-1008, ¶61.) A natural language response 78 is generated from the response representation. (EX-1008, ¶¶61, 63, Figure 4.)



Ross, Figure 4

Accordingly, the Barbara-Ross-Kellner system discloses “*generating [by the computer system] a response to the natural language utterance based on the interpretation.*”

The Barbara-Ross-Kellner combination also discloses “*generating [by the computer system] a response to the natural language utterance based on ... the identified manner in which the natural language utterance was spoken.*” For example, Kellner discloses “a user's interaction behavior is incorporated in the associated user model so that, more particularly, the output modalities used by the

dialog system during a dialog are used as well as possible in dependence on the use of various available input modalities....” (EX-1023, ¶5.) Specifically, Kellner discloses “[i]n dependence on the determined style data, the style of speech outputs is adapted, i.e. respective polite phrases and the same address are used, the speech level is adapted to the detected speech level; in dependence on the information density in a speech input more or fewer extensive system outputs are generated and the vocabulary used for the speech outputs is selected accordingly.” (EX-1023, ¶23.)

As a result, “[t]he dialog system is thus in a position **to generate system outputs adapted to a user's style of speech** with a style of speech considered pleasant by the respective user” and “the inhibition threshold of the use of the dialog system can be lowered.” (EX-1023, ¶5.)

C. Dependent Claims

1. Claims 2, 13

The Barbara-Ross-Kellner combination discloses this limitation.

As noted above (§V.B.8), Kellner identifies numerous manners in which an utterance can be spoken, including tone (“number of polite phrases”) and word use (use of “colloquial language” or “information density”). (EX-1023, ¶¶15-22.)

2. Claims 3, 14

The Barbara-Ross-Kellner combination discloses “*wherein the response comprises a voice response, and wherein generating the voice response based on the identified manner comprises varying ... word use.*”

Kellner discloses “the style of speech **outputs is adapted.**” For example, “respective polite phrases and the same address are used, the speech level is adapted to the detected speech level.” (EX-1023, ¶23.)

As a result, “[t]he dialog system is thus in a position to generate system outputs adapted to a user’s style of speech with a style of speech considered pleasant by the respective user.” (EX-1023, ¶5.)

3. Claims 4, 15

a) Contextual Signifiers And/Or Grammatical Rules [4A]/[15A]

Barbara-Ross-Kellner discloses [4A]/[15A] in two ways. (EX-1003, ¶¶254-257.)

First, Ross discloses “the language generation module 54 uses **rules 86** to choose an appropriate LG template 74 to instantiate.” (EX-1008, ¶¶77, 63, Figure 4.)

Second, Kellner discloses “[w]hen the style of speech of a user input is determined, for example evaluations with respect to [various] characterizing features” (EX-1023, ¶¶15-22.)

A POSITA would have understood that characterizing features, e.g., number of polite phrases used, way of addressing, speech level, which includes dialect, are “*grammatical rules*,” because they are used to form sentences. (EX-1003, ¶257.) Accordingly, the Barbara-Ross-Kellner combination discloses “*obtaining, by the computer system, contextual signifiers and/or grammatical rules*” [4A]/[15A].

b) Response Based On Contextual Signifiers And/Or Grammatical Rules [4B]/[15B]

Barbara-Ross-Kellner discloses [4B]/[15B] in two ways.

First, Ross discloses “[a]n example of the generation of a response 78 ... shows the rule 86-1 for choosing the LG syntax template 74.” (EX-1008, ¶¶63, 77, Figure 4.)

Second, Kellner discloses “[i]n dependence on the determined style data, **the style of speech outputs is adapted.**” (EX-1023, ¶23.) That is, Kellner discloses adapting the responsive sentences using the “*grammatical rules*” e.g., polite phrases and a specific form of address are used, or the speech level. Therefore, the system “*cooperates with the user*” by adapting to the user’s manner and to “a style of speech considered pleasant by the respective user.” (EX-1023, ¶5.)

Accordingly, the Barbara-Ross-Kellner combination discloses “*wherein generating the response based on the identified manner in which the natural language utterance was spoken comprises using the obtained contextual signifiers*

and/or grammatical rules to generate sentences for use as response sets to cooperate with the user” [4B]/[15B].

4. Claims 5, 16

Barbara renders obvious claims 5 and 16. (EX-1003, ¶¶261-262.)

Barbara renders obvious “*the long-term knowledge is associated with a first user*” and “*generating a profile associated with the first user based on the long-term knowledge.*” For example, “[t]he information database(s) is built up over time, **by recording personal details or preferences from what has been said, or from information directly entered into the system by the user.**” (EX-1007, 88.)” Accordingly, Barbara discloses recording personal information, e.g., details and preferences, for a user. A POSITA would have found obvious to store the personal recorded information in a profile associated with each user, so that the system can use the personal information for each user to disambiguate interpretations. (EX-1003, ¶261.) For example, as discussed in §V.B.7, based on stored user profile data, the system determines that “[s]omeone who is not involved in financial markets is unlikely to be asking about the Nikeii”, implying the system infers preference/employment information about the user. (EX-1007, ¶98.) Barbara further discloses that “the information database could in fact be stored at the user's terminal” to maintain “the privacy of the user,” (EX-1007, ¶89), further confirming that “*the long-term knowledge is associated with a first user.*”

Barbara also renders obvious that “*the context for the natural language utterance is further determined based on the profile associated with the first user.*”

For example, Barbara discloses an example where “the user asks for the date of Toto’s last flea vaccination.” (EX-1007, ¶148.) Barbara explains that “[t]he database has flea vaccination information for Toto but not for the others,” (EX-1007, ¶148) and therefore “it is much more likely that they are referring to Toto the dog, rather than Toto the restaurant or Tokyo the city.” (EX-1007, ¶148.) A POSITA would have found obvious to hold the flea vaccination information for Toto in the user profile, because Toto is the user’s dog. (EX-1003, ¶262.) Accordingly, Barbara discloses that the “*context for the natural language utterance,*” e.g., that the utterance refers to pets or canine (the score for the canine interpretation may be raised) (EX-1007, ¶148), is “*determined based on the profile associated with the first user,*” e.g., the flea vaccination information in the user’s profile.

5. Claims 6, 17

The Barbara-Ross-Kellner combination discloses “*adapting, by the computer system, the response based on a response format associated with the identified manner.*”

For example, Kellner discloses “[i]n dependence on the determined style data, **the style of speech outputs is adapted.**” (EX-1023, ¶23.) That is, Kellner

discloses “*adapting the response based on a response format,*” e.g., a format that uses respective polite phrases or a format that uses the same address, e.g., “you.”

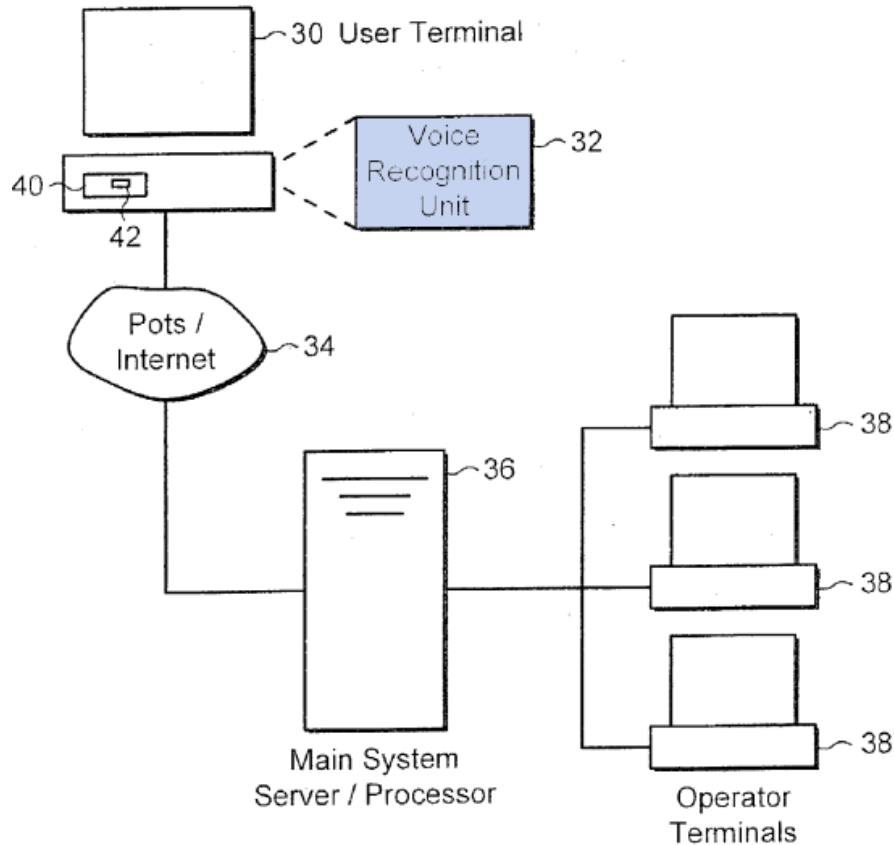
As a result, “[t]he dialog system is thus in a position to generate system outputs adapted to a user's style of speech with a style of speech considered pleasant by the respective user. As a result, the inhibition threshold of the use of the dialog system can be lowered.” (EX-1023, ¶5.)

Accordingly, the Barbara-Ross-Kellner combination discloses “*adapting, by the computer system, the response based on a response format associated with the identified manner.*”

6. Claims 7, 18

Barbara discloses “*providing, by the computer system, the natural language utterance as an input to a speech recognition engine; [7A]/[18A] and obtaining, by the computer system, the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine. [7B]/[18B].*”

Barbara’s Figure 6 shows a user terminal 30 that “has a voice interface equipment 32 [shaded blue] that is operable to **receive the spoken word and translate it into an electronic format.**” [7A]/[18A] (EX-1007, ¶83.) As explained in §V.B.2, the voice recognition unit 32 can be included in the server 36.



Barbara, Figure 6

Barbara also discloses that “[t]he end-user's machine and the voice recognition/interface service must collectively recognize that the user has said something and determine what has been said. The process of voice recognition is thus standard. The software detects when the user is speaking ...” (EX-1007, ¶86.)

In Barbara’s system, “[w]hen the application is informed by the voice recognition package that an utterance has been detected, the software asks the voice recognition engine for its recognized text phrases. **A request packet is then generated, the request comprising the best guesses of the voice recognition**

engine as to what was said, with some indications of their likelihood and also an audio file of the utterance” [7B]/[18B]. (EX-1007, ¶86.)

7. Claims 8, 19

Barbara discloses “*causing, by the computer system, the response to the natural language utterance to be provided to the user.*”

Barbara discloses that after the intended meaning has been established, “[t]he client system now performs the correct intention. **The exact way in which the system looks and responds depends on the nature of the system being operated.** Where the system is a Windows personal computer, a conventional application such as Microsoft Word may be driven via OLE Automation, for instance, **so the return intention** may be formatted as a sequence of OLE automation calls and the feedback is given by **the application visibly responding.** **The user terminal** could however be a domestic appliance, in which case, **the appliance may either speak back** or perform some action (e.g. start the spin cycle) or both. Here the response could be a series of phonemes to be spoken and a command as to the state to enter (start rinse, cook toast, start heating).” (EX-1007, ¶107.); *see also*, EX-1007, ¶¶119-162 (describing examples related to an email application), 166.)

Barbara also discloses that “[a] typical expert system works **by asking the user questions, and continuing to ask relevant questions, narrowing down the**

diagnosis until a conclusion of satisfactory confidence is reached. This is very simple to map into the framework of the system described above. To do this, standard questions are entered into the system database, together with likely answers.” (EX-1007, ¶168.)

8. Claims 9, 20

The Barbara-Ross-Kellner combination discloses “*wherein the response comprises a voice response, and wherein generating the voice response to the natural language utterance based on the identified manner comprises varying one or more of word use and/or jargon of the voice response.*”

Specifically, Kellner discloses “[i]n dependence on the determined style data, the style of speech outputs is adapted, i.e. respective polite phrases and the same address are used, the speech level is adapted to the detected speech level; in dependence on the information density in a speech input more or fewer extensive system outputs are generated and the vocabulary used for the speech outputs is selected accordingly.

That is Kellner teaches “*varying one or more of word use of the voice response,*” e.g., by using respective polite phrases, a respective way of addressing, adapting the speech level, e.g., by using a dialect or colloquial language. (See EX-1023, ¶¶15-22.)

9. Claims 10, 21

The Barbara-Ross-Kellner combination discloses “*adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use,*” for the reasons discussed in §§V.C.2 and V.C.8.

Moreover, Kellner discloses that “[t]he dialog system is thus in a position to generate system outputs adapted to a user's style of speech with **a style of speech considered pleasant by the respective user**. As a result, the inhibition threshold of the use of the dialog system can be lowered.” (EX-1023, ¶5.) That is, the system is perceived by the respective user to have a pleasant personality.

10. Claims 11, 22

The Barbara-Ross-Kellner combination discloses “*generating a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.*”

For example, the Barbara-Ross-Kellner combination discloses “*generating a response that is sensitive to context.*” As discussed in §§V.B.5 and V.B.9, the combination discloses “*generating a response*” to the natural language utterance **based on the interpretation**, and as further explained in §V.B.5, the combination

discloses determining an interpretation of the natural language utterance **based on the identified context**.

The Barbara-Ross-Kellner combination further discloses “*generating a response that is sensitive to what the user already knows about a topic.*” As discussed in §V.B.9, in connection with Barbara’s expert systems, “[a] typical expert system works by asking the user questions, and continuing to ask relevant questions, narrowing down the diagnosis until a conclusion of satisfactory confidence is reached.” (EX-1007, ¶168.) Accordingly, an exemplary “*topic*” can include a medical diagnosis, where the user provides information to the expert system during the conversation to allow the expert system to reach a diagnosis.

The Barbara-Ross-Kellner combination further discloses “*generating a response that is sensitive to short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances,*” because the combination generates a response to the natural language utterance based on the interpretation, and the interpretation is “*determine[ed] ... based on the identified context*” ([1D]/[12D]) and the “*context*” is “*identified ... based on the one or more words or phrases recognized from the natural language utterance*” as explained in §§V.B.5 and V.B.4.

In the Barbara-Ross-Kellner combination, the “*context*” is also determined based on the “*short-term knowledge and the long-term knowledge.*” For example,

Barbara explains “[c]ontextual information ... is captured as the user interacts with the system, in much the same way as interactions work in real life.” (EX-1007,

¶98.) For example,

if the user does not have a dog called Toto (or at least hasn't told the system about it), they're unlikely to be referring to it.

...

If the user has never referred to their dog before, they are likely to introduce it in some way when first using in conversation. When they do not, and the system responds incorrectly or not at all, then the user may clarify their initial request and then the contextual information may be captured. A user who asks about their dog ten times a day might be considered very likely to be referring to it.

(EX-1007, ¶98.) As a further example, if “the user asks for the date of Toto’s last flea vaccination, it is much more likely that they are referring to Toto the dog, rather than Toto the restaurant.” (EX-1007, ¶148.) That is, Barbara teaches context (pets versus restaurant) is identified from the recent dialog history which includes a reference associated with the context-domain pets (e.g., “flea vaccination”)(“*short-term knowledge*”). Barbara also teaches an utterance’s context is identified from user profile information accumulated from prior conversations, the “*long-term knowledge*” (e.g., “Toto” determined to mean “dog” due to user profile indication user has a dog named Toto.)

VI. Stipulation

To simplify the *Fintiv* analysis, Petitioners stipulate that, if IPR is instituted, Petitioners will not pursue in the related district court proceeding any ground that Petitioners raised or reasonably could have raised against the challenged claims during the instituted IPR.

VII. 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. 1-24-cv-00839-MN (D. Del., filed Jul. 18, 2024)
- *VB Assets, LLC v. SoundHound AI, Inc.*, No. 1:24-cv-01279-MN (D. Del., filed Nov. 21, 2024)
- *Samsung Electronics Co. Ltd. et al. v. VB Assets, LLC*, IPR2025-00866 (PTAB)
- *Samsung Electronics Co. Ltd. et al. v. VB Assets, LLC*, IPR2025-00868 (PTAB)

- *Samsung Electronics Co. Ltd. et al. v. VB Assets, LLC*, IPR2025-00871 (PTAB)

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	Douglas 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 K. 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.

VIII. Conclusion

IPR is respectfully requested.

Date: May 2, 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. 10,755,699 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 13,970. 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: May 2, 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. 10,755,699 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

VBA-Samsung@alston.com

Attorneys for VB Assets LLC

Date: May 2, 2025

/Lori A. Gordon/

Lori A. Gordon
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.1] A computer-implemented method of generating natural language system responses adapted based on a user's manner of speaking,

[1P.2] the method being implemented by a computer system that includes one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:

[1A] receiving, by the computer system, a user input comprising a natural language utterance;

[1B] recognizing, by the computer system, one or more words or phrases from the natural language utterance;

[1C] identifying, by the computer system, a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;

[1D] determining, by the computer system, an interpretation of the natural language utterance based on the identified context;

[1E.1] accumulating, by the computer system, short-term knowledge based on one or more natural language utterances received during a predetermined time period,

[1E.2] wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;

[1F] accumulating, by the computer system, long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;

[1G] identifying, by the computer system, a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and

[1H] generating, by the computer system, a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.

[2] The method of claim 1, wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.

[3] The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response based on the identified manner comprises varying one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.

[4A] The method of claim 1, the method further comprising: obtaining, by the computer system, contextual signifiers and/or grammatical rules,

[4B] wherein generating the response based on the identified manner in which the natural language utterance was spoken comprises using the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.

[5] The method of claim 1, wherein the long-term knowledge is associated with a first user, the method further comprising: generating, by the computer system, a profile associated with the first user based on the long-term knowledge, wherein the context for the natural language utterance is determined based further on the profile associated with the first user.

[6] The method of claim 1, wherein generating the response based on the identified manner comprises: adapting, by the computer system, the response based on a response format associated with the identified manner.

[7A] The method of claim 1, wherein recognizing the one or more words or phrases from the natural language utterance comprises: providing, by the computer system, the natural language utterance as an input to a speech recognition engine; and

[7B] obtaining, by the computer system, the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.

[8] The method of claim 1, the method further comprising: causing, by the computer system, the response to the natural language utterance to be provided to the user.

[9] The method of claim 1, wherein the response comprises a voice response, and wherein generating the voice response to the natural language utterance based on the identified manner comprises varying one or more of word use and/or jargon of the voice response.

[10] The method of claim 1, the method further comprising: adapting the response, by the computer system, to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.

[11] The method of claim 1, the method further comprising generating a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.

[12P] A system for generating natural language system responses adapted based on a user's manner of speaking, the system comprising:

[12A] one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to:

[12B] receive a user input comprising a natural language utterance;

[12C] recognize one or more words or phrases from the natural language utterance;

[12D] identify a context for the natural language utterance based on the one or more words or phrases recognized from the natural language utterance;

[12E] determine an interpretation of the natural language utterance based on the identified context;

[12F.1] accumulate short-term knowledge based on one or more natural language utterances received during a predetermined time period,

[12F.2] wherein the one or more natural language utterances received during the predetermined time period are related to a single conversation between a user and the computer system;

[12G] accumulate long-term knowledge, wherein the long-term knowledge is accumulated based on one or more natural language utterances received prior to the predetermined time period;

[12H] identify a manner in which the natural language utterance was spoken based on the short-term knowledge and the long-term knowledge; and

[12I] generate a response to the natural language utterance based on the interpretation and the identified manner in which the natural language utterance was spoken.

[13] The system of claim 12, wherein the manner in which the natural language utterance was spoken includes an indication of at least one of tone, pace, timing, inflection, word use, and/or jargon.

[14] The system of claim 12, wherein the response comprises a voice response, and wherein to generate the voice response based on the identified

manner, the one or more physical processors are further configured to: vary one or more of tone, pace, timing, inflection, word use, and/or jargon of the voice response.

[15A] The system of claim 12, wherein the one or more physical processors are further configured to: obtain contextual signifiers and/or grammatical rules,

[15B] wherein to generate the response based on the identified manner in which the natural language utterance was spoken, the one or more physical processors are further configured to use the obtained contextual signifiers and/or grammatical rules to generate sentences for use as response sets to cooperate with the user.

[16] The system of claim 12, wherein the long-term knowledge is associated with a first user, and wherein the one or more physical processors are further configured to: generate a profile associated with the first user based on the long-term knowledge, wherein the context for the natural language utterance is determined based further on the profile associated with the first user.

[17] The system of claim 12, wherein to generate the response based on the identified manner, the one or more physical processors are configured to: adapt the response based on a response format associated with the identified manner.

[18A] The system of claim 12, wherein to recognize the one or more words or phrases from the natural language utterance, the one or more physical processors are configured to: provide the natural language utterance as an input to a speech recognition engine; and

[18B] obtain the one or more words or phrases recognized from the natural language utterance as an output of the speech recognition engine.

[19] The system of claim 12, wherein the one or more physical processors are further configured to: cause the response to the natural language utterance to be provided to the user.

[20] The system of claim 12, wherein the response comprises a voice response, and wherein to generate the voice response to the natural language utterance based on the identified manner, the one or more physical processors are further configured to: vary one or more of word use and / or jargon of the voice response.

[21] The system of claim 12, wherein the one or more physical processors are further configured to: adapt the response to model a conversation, wherein adapting the response to model a conversation comprises adapting the response to have a personality by varying word use.

[22] The system of claim 12, wherein the one or more physical processors are further configured to generate a response that is sensitive to context, what the user already knows about a topic, short-term knowledge and long-term knowledge of user preferences, and words uttered by the user in one or more prior natural language utterances.