

**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

---

IPR2025-00871

U.S. Patent No. 10,510,341

---

**PETITION FOR *INTER PARTES* REVIEW OF  
U.S. PATENT NO. 10,510,341**

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

## TABLE OF CONTENTS

|      |                                                                                       |    |
|------|---------------------------------------------------------------------------------------|----|
| I.   | Grounds for Standing.....                                                             | 1  |
| II.  | Identification of Challenge .....                                                     | 1  |
|      | A.    Prior Art.....                                                                  | 1  |
|      | B.    Grounds For Challenge .....                                                     | 2  |
| III. | The '341 Patent.....                                                                  | 2  |
|      | A.    Patent Overview .....                                                           | 2  |
|      | B.    Level of Ordinary Skill .....                                                   | 3  |
|      | C.    Claim Construction.....                                                         | 4  |
| IV.  | GROUND 1: SmartKom-Kobsa Combination Renders Claims 1-18<br>Obvious.....              | 4  |
|      | A.    Combination Overview .....                                                      | 4  |
|      | 1.    SmartKom .....                                                                  | 4  |
|      | 2.    Kobsa.....                                                                      | 6  |
|      | 3.    Motivation to Combine .....                                                     | 7  |
|      | B.    Independent Claims 1, 10.....                                                   | 9  |
|      | 1.    Preambles [1P.1]-[1P.2]/[10P]/[10A] .....                                       | 9  |
|      | a)    System and Method for “Facilitating Natural<br>Language System Responses” ..... | 9  |
|      | b)    “Computer System” .....                                                         | 11 |
|      | 2.    Short-Term and Long-Term Knowledge Limitations.....                             | 12 |
|      | a)    Accumulating Short Term Knowledge [1A]/[10B].....                               | 13 |
|      | b)    Expiring Short-Short Term Knowledge [1B]/[10C] .....                            | 18 |

|    |                                                                          |    |
|----|--------------------------------------------------------------------------|----|
| c) | Accumulating Long Term Knowledge [1C.1]-<br>[1C.2]/[10D.1]-[10D.2] ..... | 19 |
| 3. | Receiving An Utterance [1D]/[10E] .....                                  | 22 |
| 4. | Determining Context [1E]/[10F].....                                      | 24 |
| 5. | Determining An Interpretation [1F]/[10G].....                            | 33 |
| 6. | Generating A Response [1G]/[10H] .....                                   | 38 |
| C. | Dependent Claims .....                                                   | 41 |
| 1. | Claims 2, 11 .....                                                       | 41 |
| 2. | Claims 3, 12 .....                                                       | 42 |
| 3. | Claims 4, 13 .....                                                       | 43 |
| 4. | Claims 5-6, 14-15.....                                                   | 46 |
| 5. | Claims 7, 16 .....                                                       | 49 |
| 6. | Claims 8, 17 .....                                                       | 50 |
| 7. | Claims 9, 18 .....                                                       | 51 |
| V. | GROUND 2: Barbara-Ross Combination Renders Obvious Claims 1-<br>18 ..... | 53 |
| A. | Combination Overview .....                                               | 53 |
| 1. | Barbara .....                                                            | 53 |
| 2. | Ross .....                                                               | 55 |
| 3. | Motivation to Combine .....                                              | 57 |
| B. | Independent Claims .....                                                 | 59 |
| 1. | Preambles [1P.1]/[1P.2]/[10P]/[10A] .....                                | 59 |
| 2. | Short-Term Knowledge [1A]/[10B] .....                                    | 60 |
| 3. | Expiring Short-Term Knowledge Items [1B]/[1C].....                       | 62 |

|      |                                                                   |    |
|------|-------------------------------------------------------------------|----|
| 4.   | Long-Term Knowledge [1C.1]/[10D.1].....                           | 62 |
| a)   | Expired Items In Long-Term Knowledge<br>[1C.2]/[10D.2] .....      | 63 |
| 5.   | Receiving An Utterance [1D]/[10E] .....                           | 64 |
| 6.   | Determining A Context [1E]/[10F].....                             | 66 |
| 7.   | Determining An Interpretation [1F]/[10G] .....                    | 67 |
| 8.   | Generating A Response [1G]/[10H] .....                            | 70 |
| C.   | Dependent Claims .....                                            | 71 |
| 1.   | Claims 2, 11 .....                                                | 71 |
| 2.   | Claims 3, 12 .....                                                | 72 |
| 3.   | Claims 4, 13 .....                                                | 73 |
| a)   | Receiving a Second Natural Language Utterance<br>[4A]/[13A] ..... | 73 |
| b)   | Identifying User Preferences .....                                | 74 |
| c)   | Updating Long-Term Knowledge .....                                | 75 |
| 4.   | Claims 5, 14 .....                                                | 76 |
| 5.   | Claims 6, 15 .....                                                | 78 |
| 6.   | Claims 7, 16 .....                                                | 78 |
| 7.   | Claims 8, 17 .....                                                | 80 |
| 8.   | Claims 9, 18 .....                                                | 81 |
| VI.  | Stipulation.....                                                  | 82 |
| VII. | Mandatory Notices.....                                            | 82 |
| A.   | Real Party In Interest.....                                       | 82 |
| B.   | Related Matters.....                                              | 82 |

|       |                                                |    |
|-------|------------------------------------------------|----|
| C.    | Notice of Counsel and Service Information..... | 83 |
| VIII. | Conclusion .....                               | 84 |

**EXHIBIT LIST**

| <b>Exh.</b> | <b>Reference</b>                                                                                                                                                                         |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001        | U.S. Patent 10,510,341 to Baldwin et al. (“the ’341 patent”)                                                                                                                             |
| 1002        | Prosecution History for U.S. Patent 10,510,341                                                                                                                                           |
| 1003        | Declaration of Stuart Lipoff in Support of <i>Inter Partes</i> Review of U.S. Patent 10,510,341                                                                                          |
| 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 10,510,341                                                                                           |
| 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 | <b><i>Reserved</i></b>                                                                                                                                                                     |
| 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”) |

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

(“Petitioners”) petition for *inter partes* review of claims 1-18 of U.S. Patent 10,510,341 (EX-1001).

## **I. Grounds for Standing**

Petitioners certify the ’341 patent is available for IPR; Petitioners are not barred or estopped.

## **II. Identification of Challenge**

### **A. Prior Art**

The ’341 patent claims priority through a series of continuations/divisionals to U.S. Patent 8,073,681, 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-1009, ¶¶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-1009, ¶¶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).

**B. Grounds For Challenge**

| Ground |     | Claims | Prior Art      |
|--------|-----|--------|----------------|
| 1      | 103 | 1-18   | SmartKom+Kobsa |
| 2      | 103 | 1-18   | Barbara+Ross   |

**III. The '341 Patent**

**A. Patent Overview**

The '341 patent discloses “a cooperative conversational voice user interface.” (EX-1001, 7:40-42.) The system receives an utterance (input 105), which is processed by speech recognition engine 110 to generate preliminary interpretations of the utterance. (EX-1001, 7:57-61, Figure 1(below).) The preliminary interpretations are provided to conversational speech engine 115 for processing. (EX-1001, 8:3-6). Conversational speech engine 115 communicates with databases 130 to generate a response. (EX-1001, 8:9-12.)

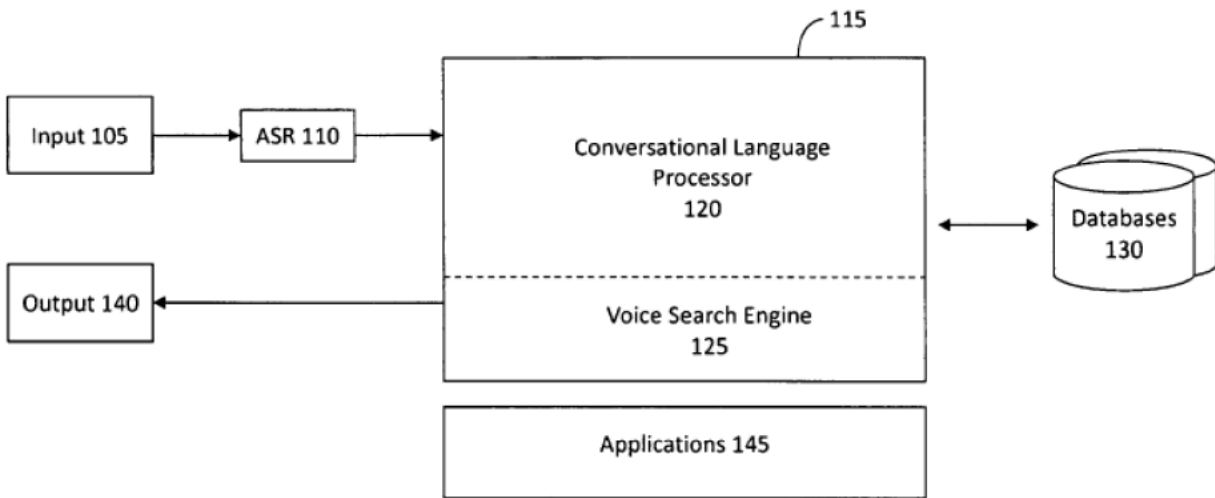


**Figure 1**

**'341 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.<sup>1</sup>

---

<sup>1</sup> The Board adopted this definition, which Patent Owner (“PO”) did not dispute, in IPR2020-01367, challenging U.S. Patent No. 8,073,681, a parent of the ’341 Patent. (EX-1010, 8.)

### C. Claim Construction

Petitioners believe no term requires construction to resolve patentability in this proceeding. Patents in the '341 Patent family were asserted in *VB Assets, LLC v. Amazon.com Servs. LLC*, No. 1:19-cv-01410-MN (D. Del., filed July 29, 2019) (“prior Amazon litigation”). While the '341 Patent was not asserted in the prior Amazon Litigation, the Court determined the terms “*short-term knowledge*,” “*long-term knowledge*,” and “*context*,” which also appear in the '341 patent’s claims, did not require construction. (See EX-1011, 2.)

## IV. GROUND 1: SmartKom-Kobsa Combination Renders Claims 1-18<sup>3</sup> 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, ¶¶106-122 (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

---

<sup>2</sup> Claim language is indicated by italics throughout.

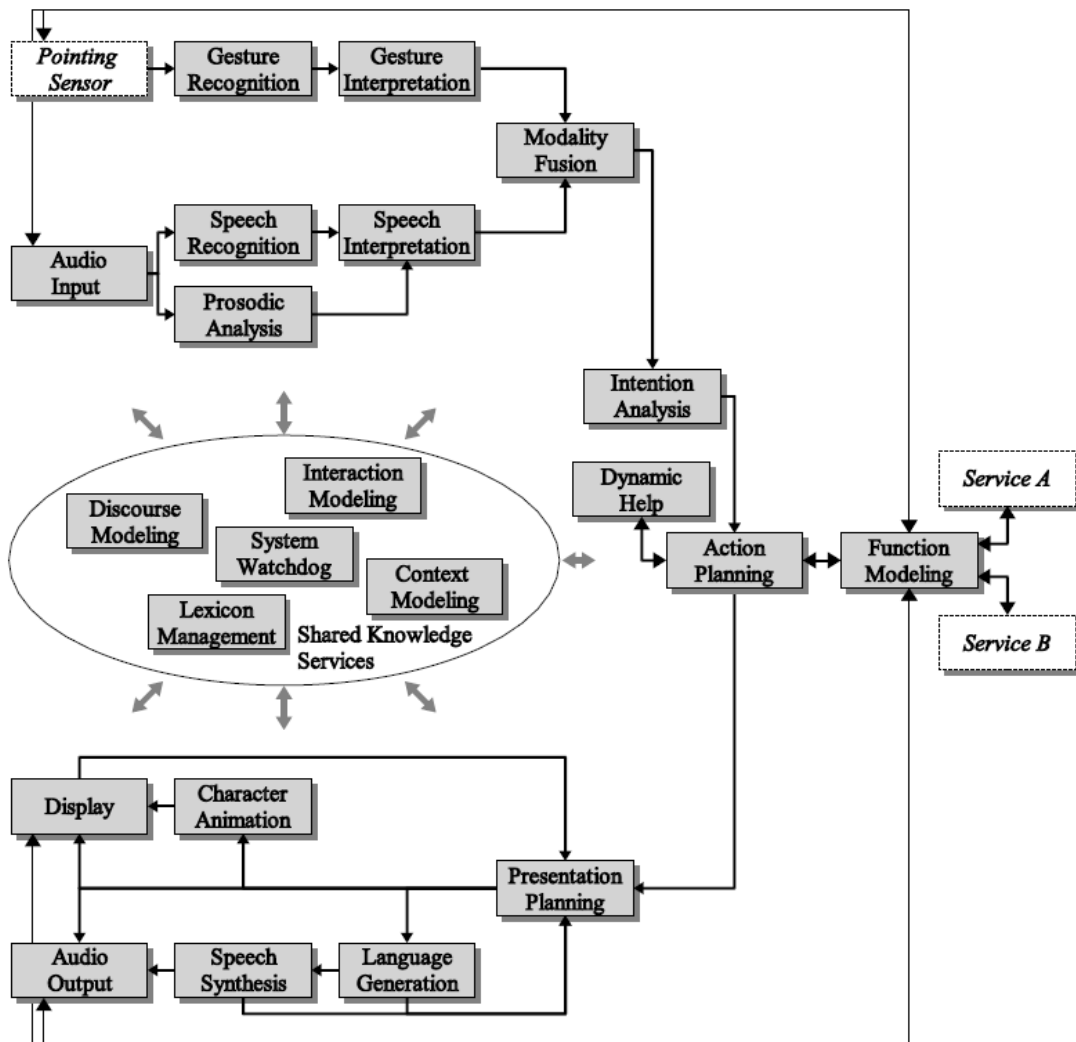
<sup>3</sup> A listing of challenged claims is provided in the Appendix.

components, each represent a “processing unit[] of the executing system.” (EX-1005, 59, Figure 2 (60)<sup>4</sup>(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.)

---

<sup>4</sup> Because SmartKom is a series of papers, Figure numbers are reused.

Petitioner therefore 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

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 ’341 patent—speech recognition systems. (EX-1001, 1:35-36 (“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), 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.)

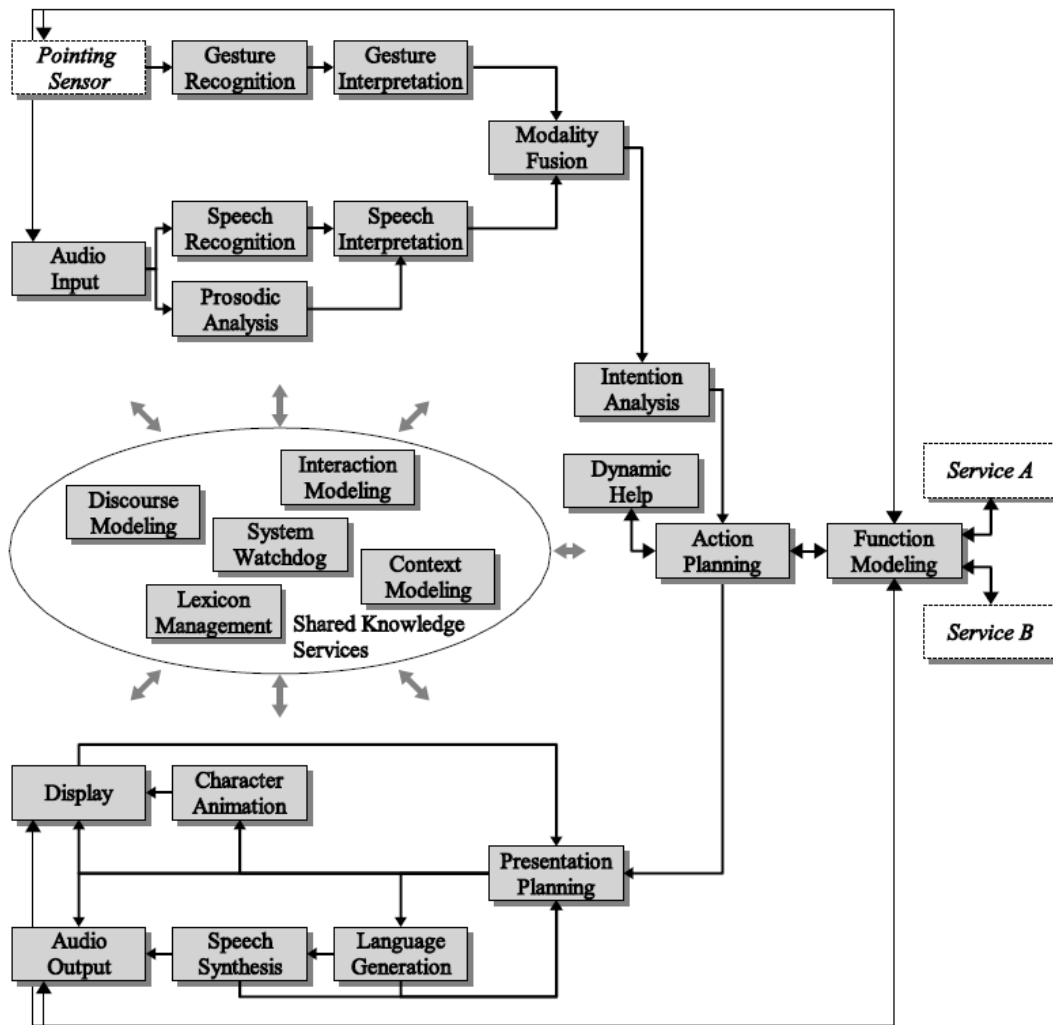
**B. Independent Claims 1, 10**

**1. Preambles [1P.1]-[1P.2]/[10P]/[10A]**

**a) System and Method for “Facilitating Natural Language System Responses”**

SmartKom discloses a “*system*” [10P], 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.



**SmartKom, Figure 2(60)**

The SmartKom system and method “*facilitat[e] natural language system responses.*” (EX-1005, 395.) 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 that is then executed during speech output.” (*Id.*) As discussed in §IV.B.6, the system also “*utiliz[es] accumulated short-term and long-term knowledge*” in generating responses.

**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]/[10A] “*perform the method*”[1P.2] or “*configure the one or more physical processors* [10A].”

SmartKom describes implementations that execute on a server having a main memory and 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*”<sup>5</sup> executed by “*one or more physical processors*”

---

<sup>5</sup> The ’341 patent does not use the term “*instructions*.” At most, it mentions a “conversational language **processor**.” (*See* EX-1001, 8:6-9.)

[1P.2]/[10A]. (EX-1003, ¶¶133-134, *citing* EX-1013, 489 (software is “Computer program; instructions that make hardware work”).)

SmartKom discloses [1P.2] and limitation [10A].

## 2. Short-Term and Long-Term Knowledge Limitations

It was well-known, prior to the '341 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.” (EX-1020, 124.) 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.)

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, 277-78, 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, 273-74.)

**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      |

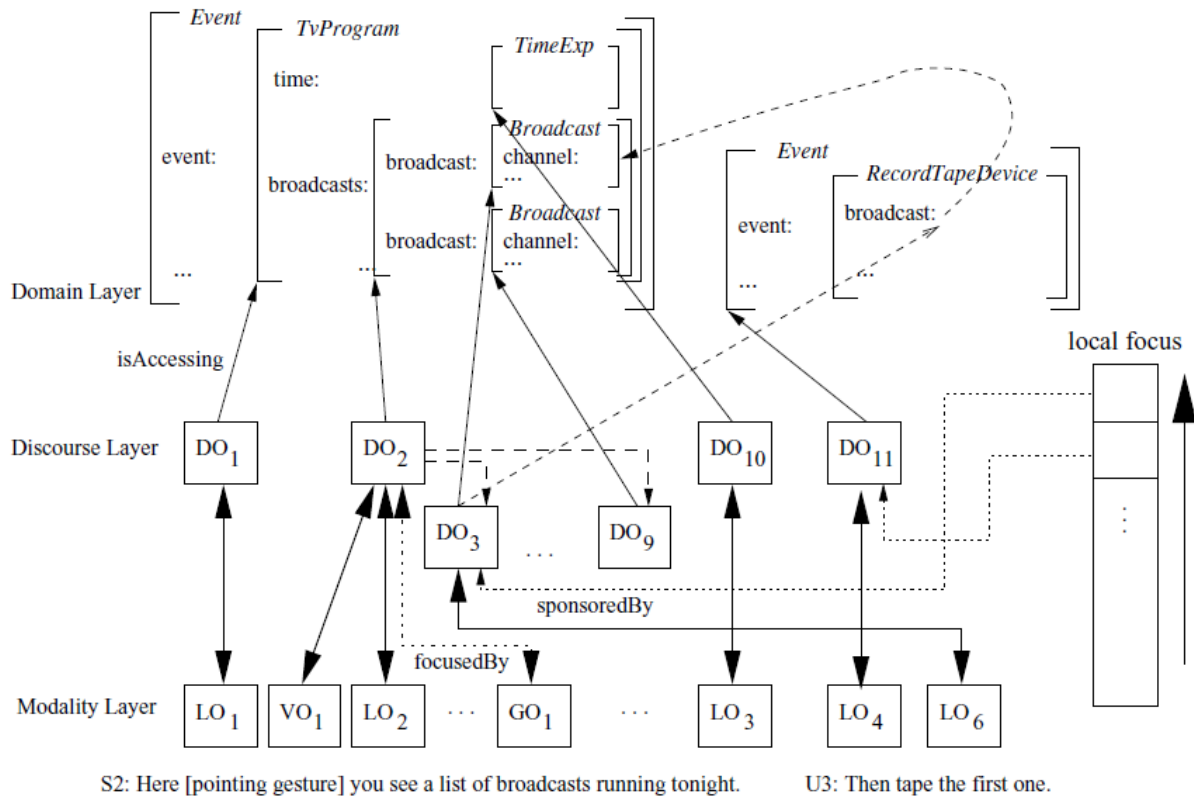
**SmartKom, Table 1(274)**

**a) Accumulating Short Term Knowledge [1A]/[10B]**

The SmartKom-Kobsa combination discloses the “*computer system*”  
“*accumulate[es].. short-term knowledge based on one or more natural language utterances...*” [1A]/[10B].

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*” that is “*based on one or more [received] natural language utterances,*” consistent with the ’341 patent’s disclosure. (EX-1003, ¶¶136, 142; EX-1001, 5:17-19 (short-term knowledge includes “input received during a single conversation”).) In SmartKom, the dialog history is referred to as the “discourse state” and is stored in discourse memory. (EX-1005, 239, 242, 246.)

SmartKom accumulates dialog (discourse state) in discourse memory and uses this knowledge to enrich and score intention hypotheses (representations of the possible interpretations of an utterance). (See EX-1005, 237; EX-1003, ¶¶114-115, 142-145.) 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 state “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, ¶145.)

The situation model, which records “time, place” of an utterance is also a type of “*short-term knowledge*” is based on the received utterance (i.e., the time and location of the utterance), consistent with the ’341 patent’s disclosure. (EX-1003, ¶¶136, 146; EX-1001, 5:19-24 (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 a type of “*short-term knowledge*” based on received utterances. (EX-1003, ¶147.)

SmartKom accumulates situational and user-state knowledge during a conversation. (EX-1003, ¶¶148-149.) 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*,” consistent with the ’341 patent. Regarding “*predetermined time period*,” the claims recite that the “*short-term knowledge*” is accumulated “*during a predetermined time period*” and “*items of short-term knowledge*” are expired “*prior to the predetermined time period*.” The ’341 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, 14:38-45.)

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 period of time.*”

**b) Expiring Short-Short Term Knowledge [1B]/[10C]**

The SmartKom-Kobsa combination discloses a “*computer system*”/ “*physical processor(s)*” that expires “*one or more items of short-term knowledge that are based on one or more natural language utterances received prior to the predetermined time period*” [1B]/[10C].

As discussed in §IV.B.2.a (above), the SmartKom-Kobsa system accumulates “*short-term knowledge based on one or more natural language utterances during a predetermined time period*” associated with a current conversation/dialog between the user and the system. As stressed by Kobsa, the

system is configured to handle multiple conversations between the user and system occurring over time. For example, the user may engage in a conversation with the system on a first day, a second conversation on the morning of a second day, and a third conversation later on the second day. The “*predetermined time period*” is associated with a current conversation and therefore any prior conversation is “*prior to the predetermined time period.*”

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.) That is, the “*short-term knowledge*” accumulated during this prior conversation (i.e., utterances “*received prior to the predetermined time period*”) “*expir[es]*” before the current conversation and its “*predetermined time period*” (i.e., the “*short-term knowledge*” expires “*prior to the predetermined time period.*”

**c) Accumulating Long Term Knowledge [1C.1]-[1C.2]/[10D.1]-[10D.2]**

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, wherein the long-term knowledge includes at least one of the one or more expired*

*items of short-term knowledge*” [1C.1]-[1C.2]/[10D.1]-[10D.2]. (See EX-1003, ¶¶150-159 (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* EX-1006, 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, 5:42-46 (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.)

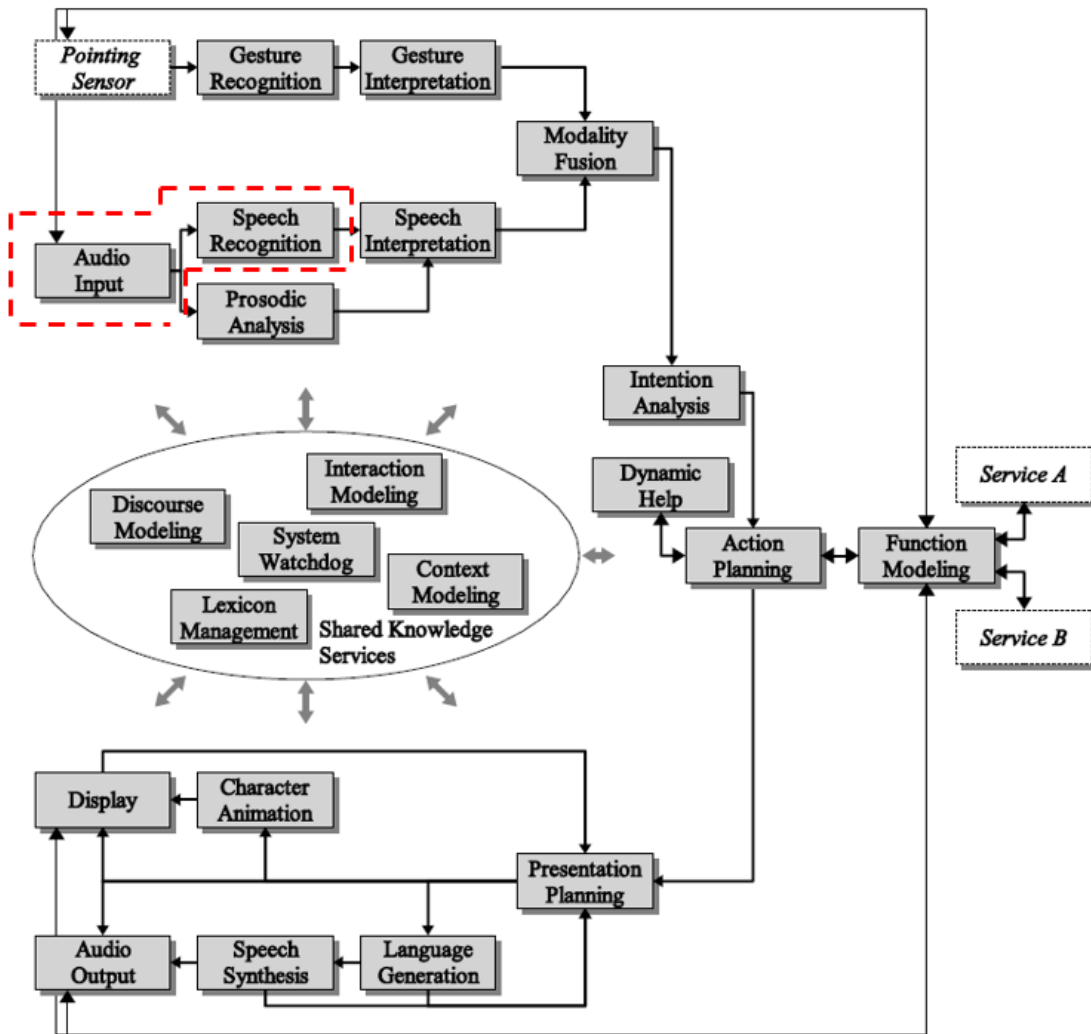
Kobsa further teaches “at the end of a dialog session,” the system “records all information about the user inferred from his/her dialog behavior in a corresponding file.” (EX-1006, 10-11.) That is, the system updates the long-term user model with information/relevant data before the session ends and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

Because the user profile information is transferred to long-term knowledge at the end of a conversation, SmartKom-Kobsa discloses “*the long-term knowledge includes at least one of the one or more expired items of short-term knowledge*” [1C.2]/[10D.2].

### **3. Receiving An Utterance [1D]/[10E]**

SmartKom discloses a “*computer system*”/ “*physical processor(s)*” that receives “*a first natural utterance via an input device*” [1D]/[10E].

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 into an orthographic transcription representing the utterance of the speaker.” (See EX-1005, 85, 212, 440, 457, 530, Figure 2(60).) The audio input component is “*an input device*” that “*receive[s] a first natural language utterance*” from a user. (EX-1005, 85.) Additionally or alternatively, because the speech recognizer receives the acoustic input voice signal, it is also an “*input device.*” (EX-1005, 85-86, Figure 1(86).)



SmartKom, Figure 2(60)

#### 4. Determining Context [1E]/[10F]

The SmartKom-Kobsa combination discloses a “computer system”/ “physical processor(s)” that determines “a first context for the first natural language utterance based on the short-term knowledge and the long-term knowledge” [1E]/[10F]. (EX-1003, ¶¶160-172.) 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 shared knowledge*”) and context-domain information from the domain model (“*long-term shared knowledge*”). (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.)

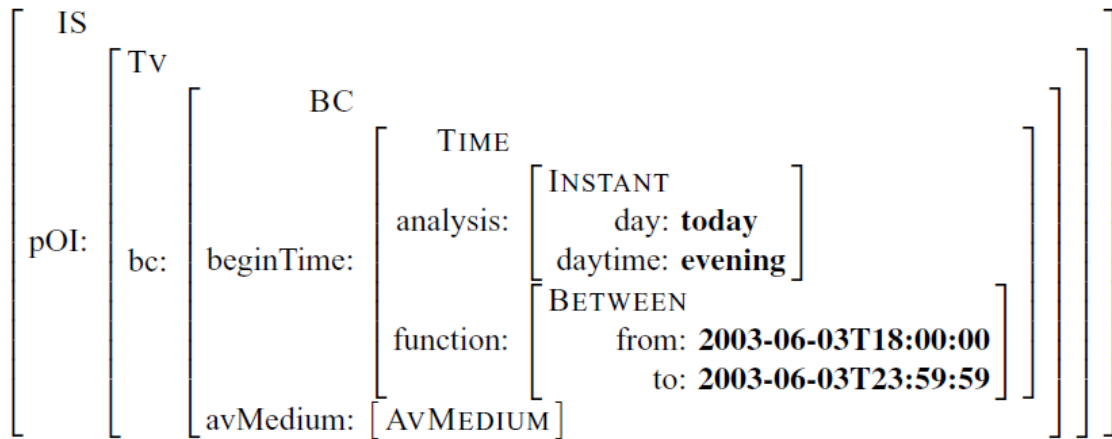


**Fig. 2.** TFS representation for the utterance *What's on TV tonight?* (*IS* = InformationSearch, *pOI* = pieceOfInformation, *Tv* = TvProgram, *bc* = broadcast)

### 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., determines a context. (EX-1003, ¶¶193-95.)

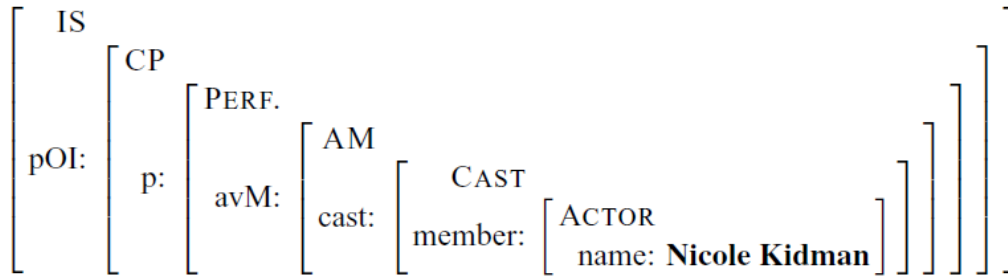


**Fig. 3.** One possible interpretation of the utterance *Is there a movie with Nicole Kidman?* (*IS* = InformationSearch, *pOI* = pieceOfInformation, *CP* = CinemaProgram, *p* = performance, *P* = Performance, *bc* = Broadcast, *Bc* = Broadcast)

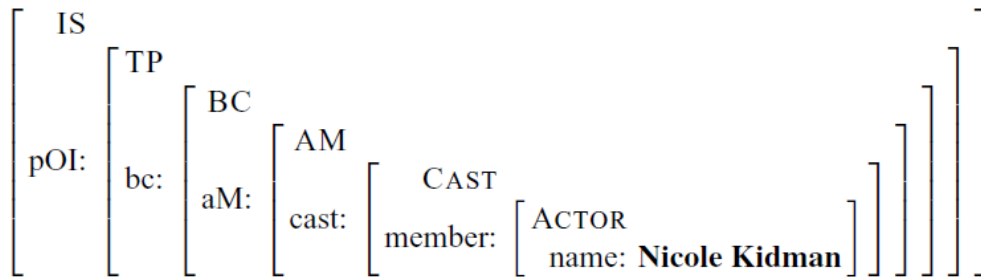


**Fig. 4.** A second possible interpretation of the utterance *Is there a movie with Nicole Kidman?* (*IS* = InformationSearch, *pOI* = pieceOfInformation, *CP* = CinemaProgram, *p* = performance, *P* = Performance, *aM* = avMedium, *AM* = AvMedium)

### 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. (EX-1005, 266.) 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, ¶195.)

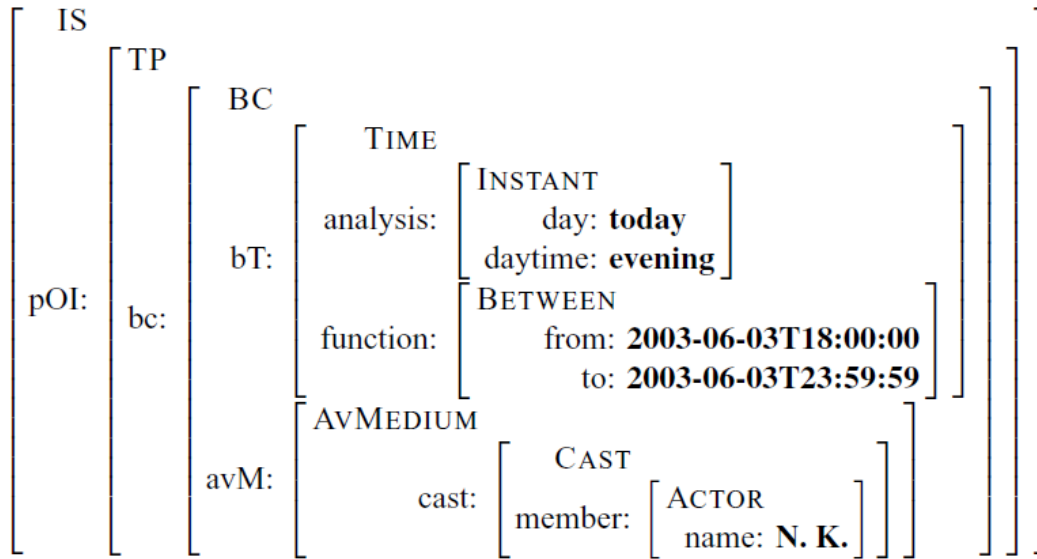


**Fig. 5.** Interpretation of the utterance *Is there a movie with Nicole Kidman?* with respect to the previous discourse context

### 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 estimation 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) which is part of the “*short-term shared knowledge*” is added. (EX-1003, ¶200.) 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 “*long-term shared knowledge*”

from the user model such as the user’s preferences or expertise in the identified context-domain. (See §IV.B.4; EX-1005, 274.) These enrichments assist with interpretation of user intent as well as shaping response generation. (§IV.B.6.)

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

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

**SmartKom, Table 2(277)**

Thus, the context modeler of the SmartKom-Kobsa combination identifies a context based on dynamic situational knowledge (“*short-term shared knowledge*”), dynamic instances (e.g., position of user when providing current utterance; “*short-term shared knowledge*”), and common ground knowledge (“*long-term shared-*

*knowledge*”), and user-specific profile information (e.g., experience, preferences, etc.; “*long-term shared-knowledge*”) and incorporates this context with the discourse and context-domain context (e.g., television, cinema program, etc.) in each hypothesis. (EX-1003, ¶¶166-171; *see, e.g.*, EX-1005, 364 (“SmartKom domains such as electronic program guides (EPG) for TV’s, cinema programs, and movie information”), 21.)

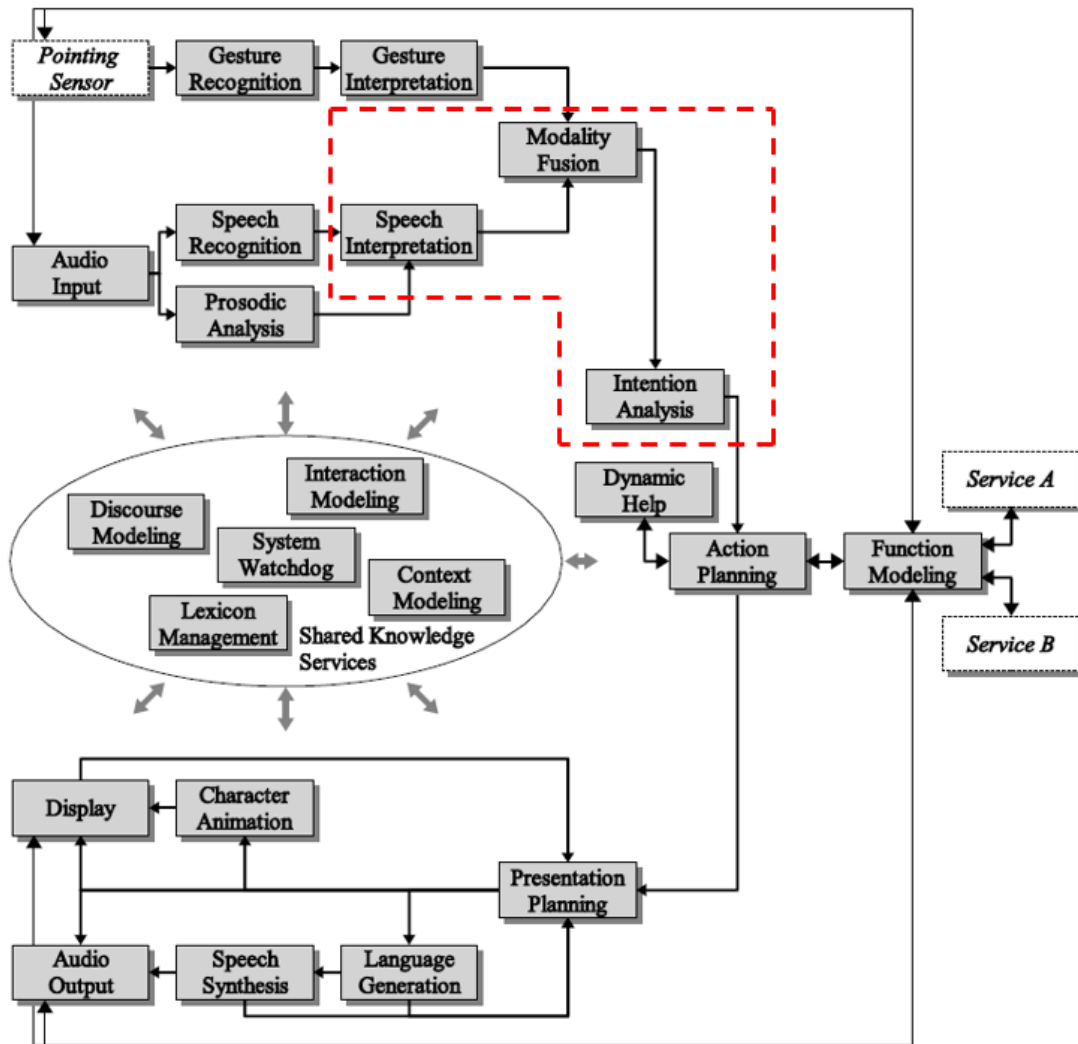
The context modeler also “compute[s] [a] contextual coherence score[.]” (EX-1005, 276, 271.) The scoring task involves using the “ontological domain context to measure the semantic coherence of the individual interpretation” and “using dynamic situational and discourse information, e.g., previous ontological 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.) Table 2(277) above includes an example of the contextual coherence score added into hypothesis of the intention lattice.

SmartKom-Kobsa combination therefore discloses “*determining [by the computer system,] ... a first context for the first natural language utterance*

[context-domain, situational context, short-term user state, long-term user context] *based on the short-term knowledge* [user state, short-term user model, situational model] *and the long-term knowledge* [long-term user model, context-domain model]” [1E]/[10F].

## **5. Determining An Interpretation [1F]/[10G]**

SmartKom discloses a “*computer system*”/ “*physical processor(s)*” that determines “*based on the first context, an interpretation of the first natural language utterance*”/ “*determine an interpretation of the first natural language utterance based on the first context.*” 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.)

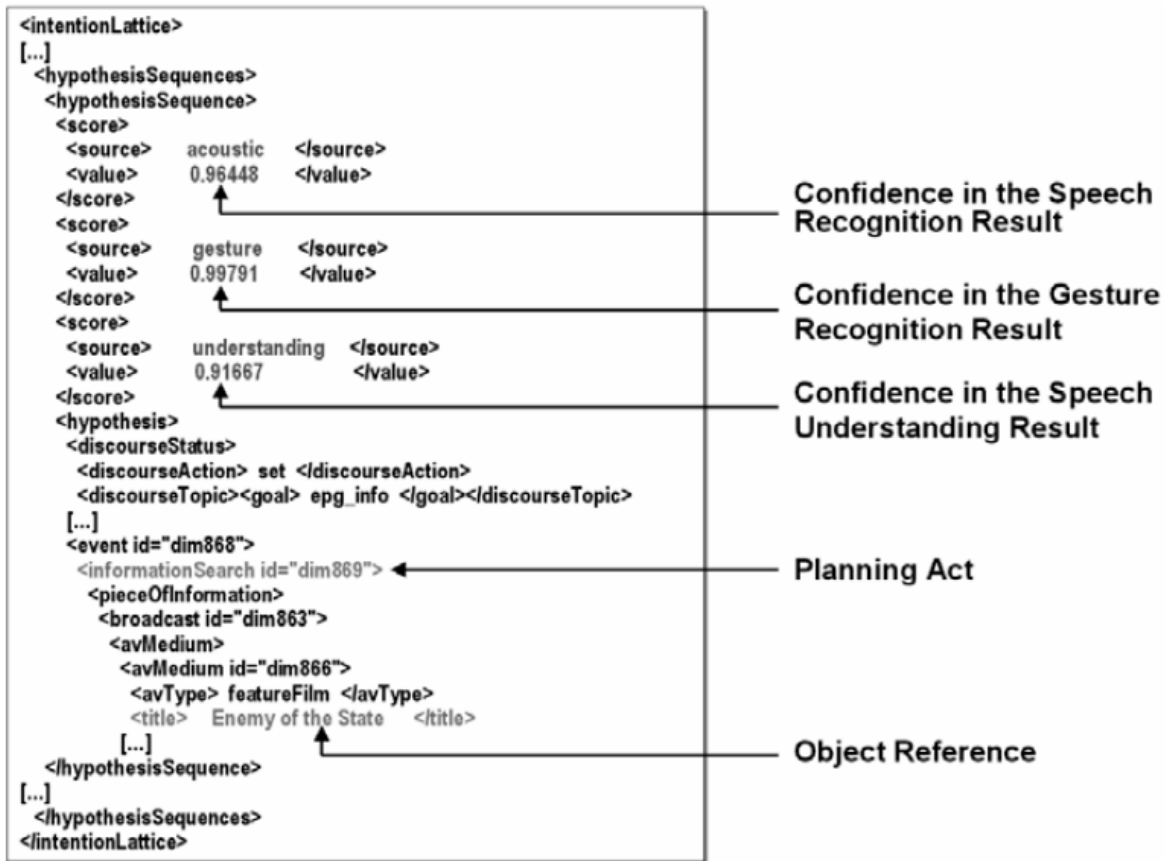


**SmartKom, Figure 2(60)**

The SmartKom modules transfer information through an “intention lattice,” containing 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.) In SmartKom, the “basic information flow from user input to system output continuously adds further processing results so that [this] representational structure [i.e., intention lattice] will be refined step-by-step.” (EX-1005, 62-63.) That is, the

intention recognizer receives an intention lattice from the modality fusion components including a set of hypotheses. (EX-1005, 62-63.)

The intention lattice shared between components 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)

In operation, 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 a corresponding score value (confidence in the speech recognition result), 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 of the first natural language utterance*” and therefore SmartKom discloses “*based on the first context, an interpretation of the first natural language utterance.*”

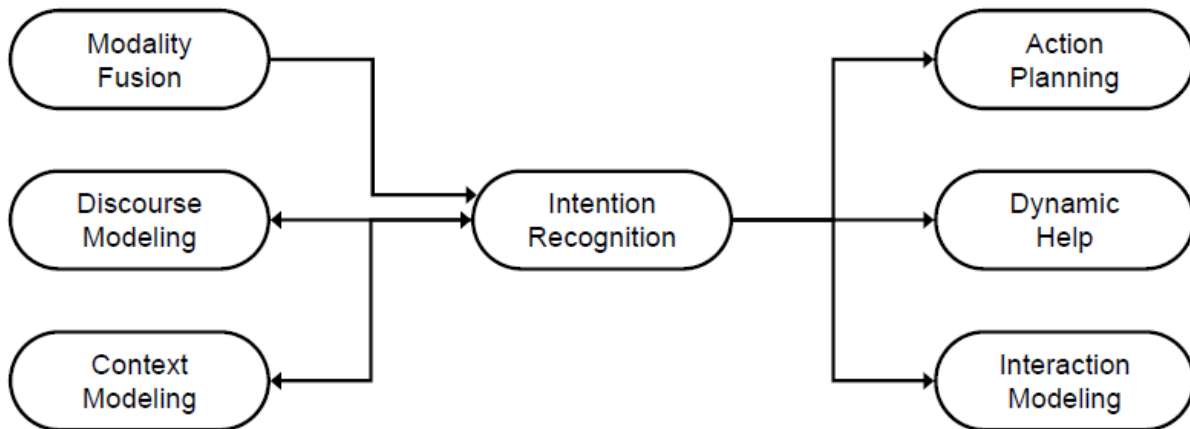


**Fig. 1.** Section of the multimodal interaction from the intention recognition point of view

**SmartKom, Figure 1(286)**

**6. Generating A Response [1G]/[10H]**

SmartKom discloses a “*computer system*”/“*physical processor(s)*” that generates “*a first response to the first natural language utterance based on the interpretation*” [1G]/[10H].

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] a first response to the first natural language utterance.*” (See EX-1005, 302, Figure 1(303)(below).) The presentation planner, natural language generation [NLG] component, and speech synthesis cooperate to present the response multi-modally to the user. (See EX-1005, 396, 401.)

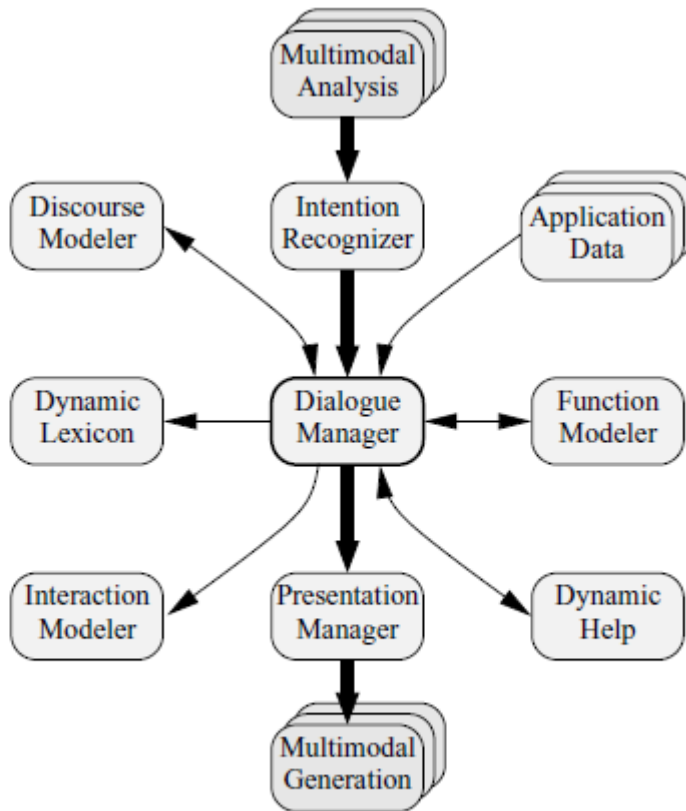


**Fig. 1.** Modules in direct communication with the dialogue manager

**SmartKom, Figure 1(303)**

The action planner “*generat[es] a first 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”). (EX-1005, Table 3(313).) 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).)

## C. Dependent Claims

### 1. Claims 2, 11

The SmartKom-Kobsa combination discloses “*the one or more natural language utterances received during the predetermined time period are related to a single conversation between the user and the computer system*” [2]/[11].

SmartKom is designed to “allow for more natural, **conversational** dialogue that is not system-driven, users can revise their input freely and often do so unexpectedly.” (EX-1005, 255-256; *see also* EX-1005, 281 (“inclusion of such contextual interpretation in natural language processing can enable natural language understanding systems to become more **conversational** without losing the reliability of restricted dialogue systems”).)

SmartKom therefore provides numerous examples of utterances occurring in a conversation between the user and the computer system. For example:

U1: *What’s on at the movies tonight?*

S1: (Displays a list of movies) *Here [↗]<sup>3</sup> you can see tonight’s cinema program.*

U2: *And what’s on TV?*

S2: (Displays a list of broadcasts) *Here [↗] are tonight’s broadcasts.*

U3: *Is there a movie with Arnold Schwarzenegger?*

(EX-1005, 238.) As discussed in §IV.B.2.a, the “*predetermined time period*” is the shorter of (1) the duration of the conversation/dialog or (2) the size of short term storage (e.g., 30 minutes). As highlighted by the above example, the user

utterances are “*related to a single conversation between the user and the computer system*” and occur “*during the predetermined time.*”

## 2. Claims 3, 12

The combination of SmartKom and Kobsa discloses “*the long-term knowledge comprises one or more of user preferences, an indication of utilized domain agents, contexts of prior natural language utterances, requests of prior natural language utterances, and tasks performed in response to prior natural language utterances*” [3]/[12].

The user model (“*long-term knowledge*”) of the SmartKom-Kobsa combination includes “*one or more user preferences.*” (EX-1006, 78, 415.) Kobsa explains that a “use of user models is for representing **user preferences.**” (EX-1006, 78.) Kobsa defines four categories of user knowledge “to be modeled” in a user model including “knowledge represent[ing] a user’s bias, preferences, and attitudes.” (EX-1006, 415.)

Kobsa provides several examples of systems using user models including user preferences including a hotel reservation system, real estate agent, and the Grundy system. (See EX-1006, 6, 13, 78, 199, 388, 414-416.) The “GRUNDY system, for instance, forms assumptions about the user's selection criteria for books to read from information about personality characteristics supplied by him/her.” (EX-1006, 13.)

SmartKom also references use of “user preferences.” In SmartKom, “personal preferences” are an “important decision in the presentation of system output in a multimodal dialogue system like SmartKom.” (EX-1005, 407.) For example, SmartKom explains that “[p]references for specific genres or actors are regarded by the presentation planning component.” (EX-1005, 336.) SmartKom discloses the system takes user preferences into account when responding to the user, explaining that “[i]f the **system knows positive or negative preferences, it first presents objects that contain a preferred feature.**” (EX-1005, 320-321.) SmartKom explains that after a preference is stored, “the system’s preference management” component “will generate the offer to present objects in the future according to the preference.” (EX-1005, 336.)

### 3. Claims 4, 13

SmartKom discloses a “*computer system*”/“*physical processor[s]*” that receives “*a second natural language utterance in response to the first response*” [4A]/[13A].

As discussed in §IV.B.6, as part of the action plan, SmartKom contemplates “*a second natural language utterance in response to the first response.*” For example, in the action plan of Table 3(313) (reproduced below), the system responds to the user’s utterance “I want to send a fax” with 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, 312-313 (Table 3).) The user than responds with the number in a responsive utterance. (See EX-1005, 313.)

**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)

SmartKom provide numerous additional examples of conversations with multiple user utterances including a user utterance responding to a system utterance. In the example in §IV.B.4, the dialog includes the following:

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.) Utterance-2 (U-2) is “*a second natural language utterance in response to the first response.*”

SmartKom-Kobsa combination discloses “*identifying, by the computer system, one or more user preferences based on the second natural language utterance*” [4B]/[13B] and “*updating, by the computer system, the long-term knowledge based on the identified user preferences*” [4C]/[13C].

As discussed in §IV.B.2.c, the user model knowledge is acquired “*either explicitly or implicitly.*” (EX-1006, 416 (emphasis in original); *see also* EX-1005, 320-321, 336.) “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.” (EX-1006, 416.) 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. (EX-1005, 416.) For example, in U-2 (“*second natural language utterance*”), the system implies a user preference for “Nicole Kidman” films. (EX-1005, 263.)

In the combination, “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.) This action updates the stored “*long-term knowledge*.” SmartKom also discloses storing the user preferences, explaining that if an utterances expresses a preference, “[t]he system asks if it should store this positive or negative preference.” (EX-1005, 336.) SmartKom explains that after a preference is stored, “the system’s preference management” component “will generate the offer to present objects in the future according to the preference.” (EX-1005, 336.)

#### **4. Claims 5-6, 14-15**

The combination of SmartKom and Kobsa discloses that “*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*” [5]/[14].

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*” [5]/[14]. (EX-1005, 414.)

The SmartKom-Kobsa combination further generates “*a profile associated with the first user based on the long-term knowledge*” [5]/[14]. 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

As discussed in §IV.B.4, the SmartKom-Kobsa combination discloses “*the first context for the first natural language utterance is further determined based on the profile associated with the first user*” [5]/[14].

For the above reasons and the reasons discussed in §IV.B.4, the SmartKom-Kobsa combination further discloses a “*computer system*”/ “*physical processor[s]*” that generates “*one or more of a **user profile**, an environmental profile, a historical profile, and/or a cognitive profile based on the long-term knowledge, wherein subsequent natural language utterances are processed based at least on the one or*

*more of the **user profile**, the environmental profile, the historical profile, and the cognitive profile” [6]/[15].*

## **5. Claims 7, 16**

SmartKom-Kobsa combination discloses a “*computer system*”/“*physical processor[s]*” that receives “*information from a remote data source wherein the long-term knowledge is based further on the information obtained from the remote data source*” in two ways.

First, in SmartKom, “general user model information is supplied via external sources, e.g., via a user’s *SmartCard*,” and “only the interaction preferences of the users are monitored actively by the system.” (EX-1005, 276 (emphasis in original).) The user’s SmartCard is a “*remote data source*.” (EX-1005, 276.) As disclosed by SmartKom, the “general user model information” for a user forms part of the “*long-term knowledge*” for the user which is updated based on “*interactions*” with the user. (EX-1005, 276.)

Second, in SmartKom, the system accesses remote databases based on the user’s request: “By means of explicating such information and providing topical and contextually adequate values, the system can retrieve appropriate information from Web sites or databases on what is currently playing in town, produce maps featuring cinema locations and then offer further assistance in navigation or reservation, for example.” (EX-1005, 278.) When the system presents the

information from the remote database, the user may indicate a preference based on what was presented, as shown in the following example:

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; *see also*, EX-1005, 320-321 (describing identifying preferences after data presented to user on display).) The reference to “*Nicole Kidman*” is an indication of a preference. (*See* §IV.C.3; EX-1006, 13.)

Thus, as illustrated above, “*information*” is received “*from a remote data source*” and “*the long-term knowledge is based further on the information obtained from the remote data source.*”

## 6. Claims 8, 17

SmartKom discloses a “*computer system*”/“*physical processor[s]*” that determines “*an interpretation of one or more recognized words of the first natural language utterance based on the first context*” [8]/[17].

SmartKom discloses a speech recognition engine that “transforms the acoustic signal into a **sequence of hypothesized words** in orthographic representation.” (EX-1005, 86, Figure 2(60).) The speech recognition engine uses “**confidence measures**” to “estimate confidences for the correctness of the words hypothesized by the recognizer.” (EX-1005, 85, 96.) The hypotheses, each with a

corresponding score value (confidence in the speech recognition result), 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, 233.) That is, up to three hypotheses, each with “*recognized words of the first natural language utterance*” are provided to the intention recognizer.

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).) A contextual coherence score for each hypothesis in the intention lattice is provided to the intention recognizer and used to select one hypothesis. (EX-1005, 285.) And as discussed in §IV.B.5, the intention recognizer in conjunction with the discourse and context modelers, “*determin[es] ... an interpretation of one or more recognized words of the first natural language utterance based on the first context.*”

## **7. Claims 9, 18**

SmartKom discloses a “*computer system*”/ “*physical processor[s]*” that causes “*the response to the first natural language utterance to be provided to a user*” [9]/[18].

Specifically, as discussed in §IV.B.6, the action planner, also referred to as the dialogue manager, “*generat[es the] response to the first natural language utterance.*” (See EX-1005, 302, Figure 1(303).) The presentation planner, natural language generation [NLG] component, and speech synthesis cooperate to present the response multi-modally to the user. (EX-1005, 396, 401.)

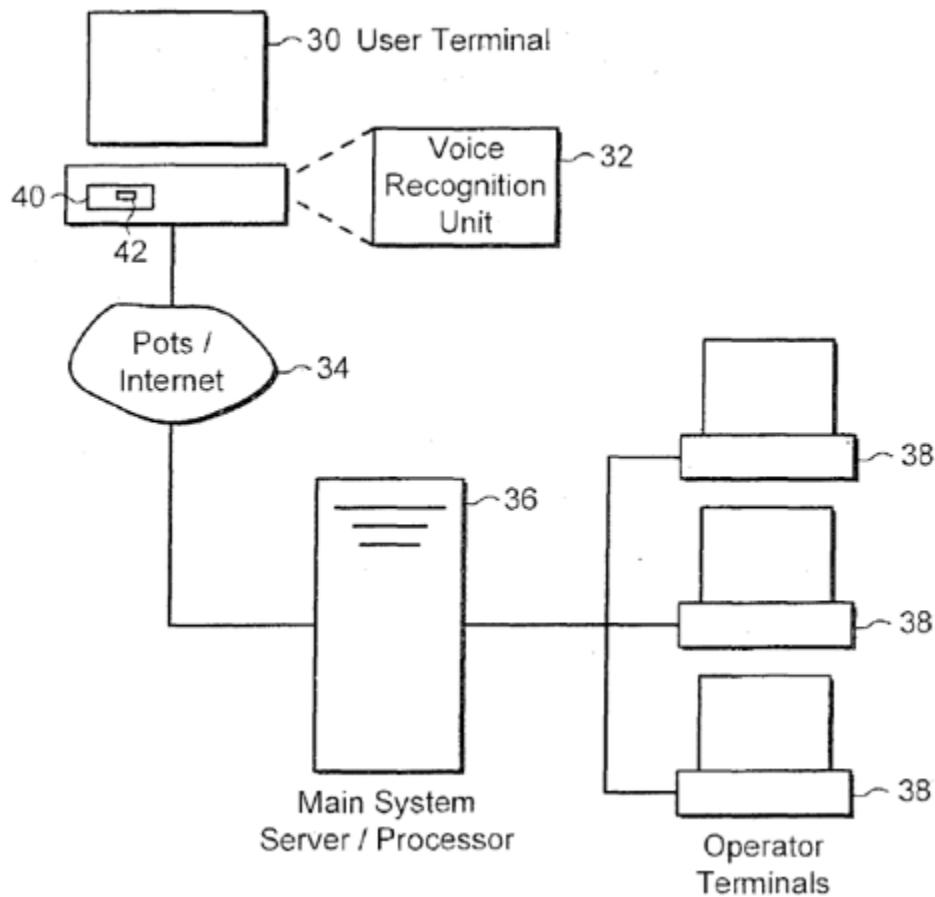
SmartKom also explains the response is adapted based on the user’s “*natural language utterance.*” SmartKom’s “[i]nteraction modeling is concerned with different aspects like available modalities and user preferences for specific forms of communication as well as the affective state of the user.” (EX-1005, 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 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.)

## **V. GROUND 2: Barbara-Ross Combination Renders Obvious Claims 1-18**

### **A. Combination Overview**

#### **1. Barbara**

Barbara's system (figure 6, below) interprets a user's spoken word (e.g., the utterance) "to determine what the user wanted, i.e. the intent of the spoken word." (EX-1007, ¶82.) The system includes user terminal 30 with voice interface equipment 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.) Contextual information is "captured as the user interacts with the system." (EX-1007, ¶98.)



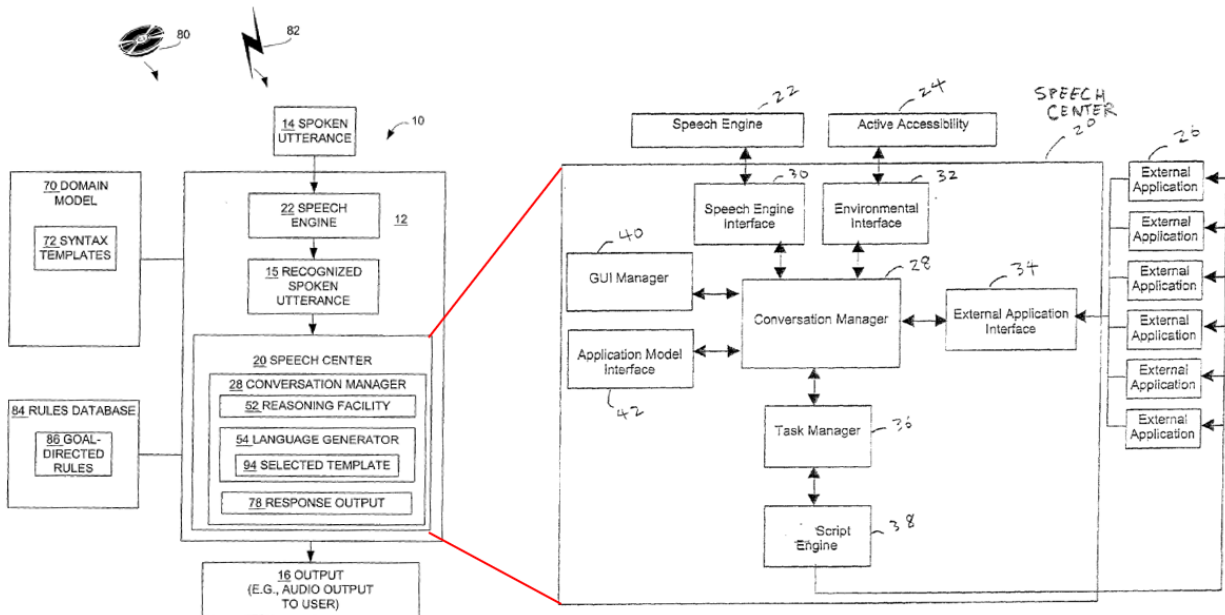
**Barbara, Figure 6**

When a user's "utterance has been detected," voice recognition engine 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 provides a response. (EX-1007, ¶¶105-107, Figure 7.)

Barbara also provides a manual correction technique when utterances "fail[] to meet the pre-determined criteria for automatic acceptance." (EX-1007, ¶¶100-104.) Notably, this operator intervention and correction "will be used heavily during" an "initial period." (EX-1007, ¶102.) (*See also*, EX-1003, ¶¶174-186.)

## 2. Ross

Ross discloses a “conversation manager [that] processes spoken utterances from a user of a computer, and develops responses to the spoken utterances.” (EX-1008, Abstract.) Ross’s system 10 comprises digital processor 12 including speech engine speech center 20 having conversational manager 28. (EX-1008, ¶¶22-24, Figures 1-2 (below).)

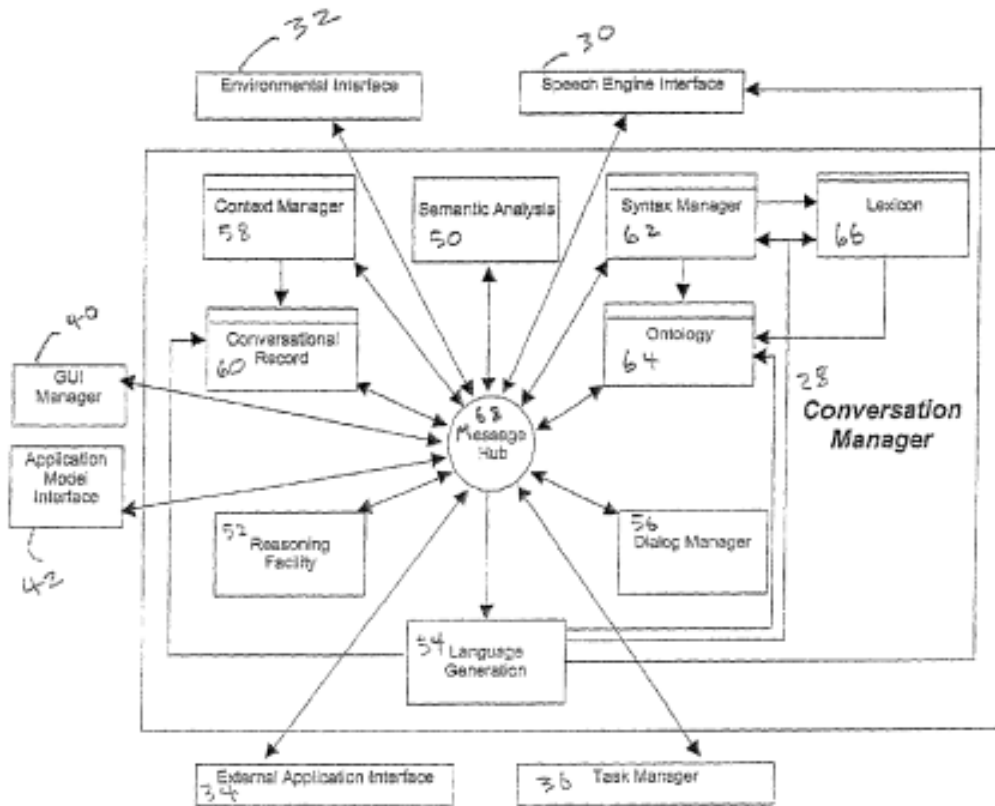


**Ross, Figure 1**

**Ross, Figure 2**

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

stores conversational record 60 including each utterance in a conversation and domain model 70. (EX-1008, ¶57.)



**Ross, Figure 3**

Ross's systems uses the conversational records and domain model to process utterances. (EX-1008, ¶22 ("a natural language response output 78 to the recognized spoken utterance 15 [is] based on" domain model 70, rules database 84, and syntax template 94).) Ross however notes that "[c]onversational speech is full of implicit and explicit references back to people and objects that were mentioned earlier." (EX-1008, ¶57.) To understand such sentences, speech center 20 "looks at

the **conversational record 60**, and finds the missing information.” (EX-1008, ¶57.) (*See also*, EX-1003, ¶¶187-199.)

### 3. Motivation to Combine

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, ¶200.)

Barbara and Ross are analogous art to the ’341 Patent. Each pertains to the same field of endeavor, i.e., speech recognition systems. (*Compare* EX-1001, 1:35-36 *with* EX-1007, ¶7 *and* EX-1008, Abstract; EX-1003, ¶201.)

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, ¶202.) 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, ¶203.) A POSITA would understand this data management approach improves storage efficiency and reduces need for extensive hardware storage. (EX-1003, ¶203.)

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, ¶204.)

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, ¶205.) 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 '341 patent. (EX-1003, ¶205.)

## B. Independent Claims

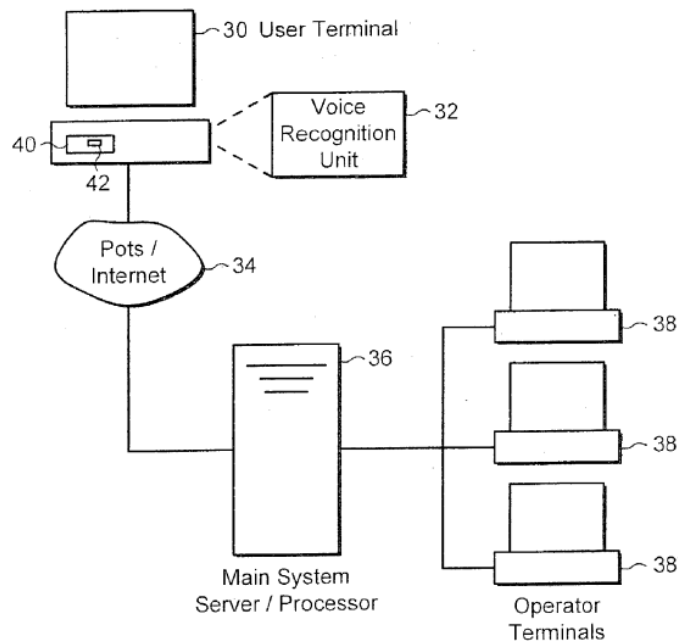
### 1. Preambles [1P.1]/[1P.2]/[10P]/[10A]

The Barbara-Ross combination discloses the preambles [1P.1]/[1P.2]/[10P]/[10A]. (EX-1003, ¶¶206-209.) Barbara discloses a “*method*” [1P.] and a “*system*” [10P]: “**system and method** for improving accuracy of signal interpretation.” (EX-1003, ¶¶206-209.)

Barbara’s system “*facilitat[es] natural language system responses.*” (EX-1003, ¶207.) For example, Barbara discloses “[t]he system can ... be 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; *see also*, EX-1007, ¶7.)

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, ¶207.) 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.)



**Barbara, Figure 6**

Barbara's system also "*utilizes accumulated short-term and long-term knowledge*" as explained in §§V.B.2, V.B.4 (below).

## **2. Short-Term Knowledge [1A]/[10B]**

In the Barbara-Ross combination, the "best guesses of the voice recognition engine as to what was said are evaluated" "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 combination uses Ross’s data organization, for example the knowledge base is short-term storage that uses Ross’s conversational record 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 combination discloses “*accumulating short-term knowledge based on one or more natural language utterances received during a predetermined time period*” [1A]/[10B]. (EX-1003, ¶¶210-214.)

### 3. Expiring Short-Term Knowledge Items [1B]/[1C]

In the Barbara-Ross system, “[e]ach utterance is indexed in the conversational record 60, along with the results of its semantic analysis. The 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*” that had been accumulated during a prior conversation—and is no longer relevant to active goals—(i.e., utterances “*received prior to the predetermined time period*”) is purged from the conversational record, i.e., “*expir[es]*.”

Thus, the Barbara-Ross combination discloses “*expiring [by the computer system] one or more items of short-term knowledge that are based on one or more natural language utterances received prior to the predetermined time period.*” (EX-1003, ¶¶215-216.)

### 4. Long-Term Knowledge [1C.1]/[10D.1]

Barbara records (“*accumulates*”) “**personal details or preferences from what has been said**” in the “information database(s).” (EX-1007, ¶88; *see also* 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 “[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.) Thus, 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, ¶¶217-218.) Because this content “is built up over time,” the accumulated information is “*long-term knowledge.*” (EX-1003, ¶219; EX-1001, 5:42-46 (“Long-term shared knowledge may **include explicit and/or implicit user preferences**”).)

The Barbara-Ross combination therefore discloses “*accumulating, [by the computer system,] long-term knowledge based on one or more natural language utterances received prior to the predetermined time period,*” [1C.1]/[10D.1].

**a) Expired Items In Long-Term Knowledge [1C.2]/[10D.2]**

Ross’s conversational record holds “[e]ach utterance ..., along with the results of its semantic analysis” that is relevant to the user’s current conversation, e.g., “relevant to active goals.” (EX-1008, ¶57.) For example, when a user engages in a conversation that introduces their dog “Toto,” the “conversational record,” e.g., *short term knowledge*, would hold the utterances of the conversation and the results of their semantic analysis, for example, that the user is a dog owner or that

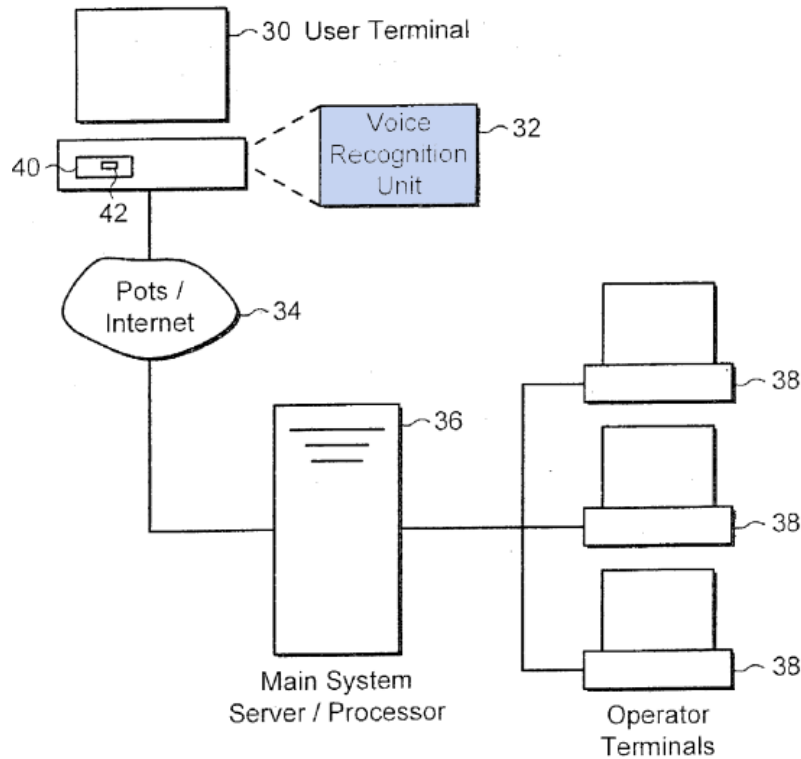
they prefer dogs over cats. After, however, “a predetermined period of time has elapsed” (EX-1008, ¶57) when the information is purged from the “conversational record,” a POSITA would have understood that, at least part of the information, e.g., “personal details or preferences from what has been said” (EX-1007, ¶88) would be moved into Barbara’s informational database, because it constitutes “[t]he history of what people have asked [that is] used to assist in scoring the possible intents” (EX-1007, ¶98). Accordingly, the Barbara and Ross combination discloses “*the long-term knowledge includes at least one of the one or more expired items of short-term knowledge*” [1C.2]/[10D.2]. (EX-1003, ¶220.)

## **5. Receiving An Utterance [1D]/[10E]**

Barbara discloses “*receiving, at the computer system, a first natural language utterance via an input device.*” [1D]/[10E].

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, Figure 6 (below).) Barbara teaches that user terminal 30 “could merely be used as a telephone, indeed it may in fact be a telephone/mobile phone, or audio recorder and the voice recognition software could be provided at a remote location.” (EX-1007, ¶83, *see also*, EX-1007, ¶86.) The voice recognition software, whether in the user device or in the server, alone or in combination with

a component in the user device (e.g., a microphone) is an “*input device*” that  
“*receive[s] a first natural utterance.*”



**Barbara, Figure 6**

In Barbara “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 what has been said. ... When 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.” (EX-1007, ¶86; *see also* EX-1007, Figure 7.)

## 6. Determining A Context [1E]/[10F]

Barbara discloses “*determining a context.*” 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 shared knowledge*”). (EX-1003, ¶¶221-224.) Barbara also teaches an utterance’s context is identified from user profile information accumulated from prior conversations, the “*long-term shared knowledge*” (e.g., “Toto” determined to

mean “dog” due to user profile indication user has a dog named Toto). (EX-1003, ¶¶225-226.)

Thus, the Barbara-Ross combination discloses “*determining [by the computer system] based on the short-term knowledge and the long-term knowledge, a first context for the first natural language utterance*” [1E]/[10F]. (EX-1003, ¶¶221-226.)

## **7. Determining An Interpretation [1F]/[10G]**

The Barbara-Ross combination discloses “*determining, by the computer system, based on the first context, an interpretation of the first natural language utterance*” [1F]/[10G].

Barbara discloses that “[t]he system can [] additionally or alternatively be used to allow text that is spoken into voice recognition equipment to be corrected or to create a voice interface where the end system responds to spoken natural language commands.” (EX-1007, ¶82.) “In the case of voice recognition the electronic version of the spoken word is checked for accuracy, corrected and then returned to the user. In the case of the voice interface, however, **the electronic version of the spoken word is checked for accuracy, corrected and interpreted to determine what the user wanted, i.e. the intent of the spoken word. Once the intent is determined**, a command signal can be sent to the user's terminal to carry out that intent. It will be appreciated that the voice interface uses the same

features as the simpler voice recognition system, but with the addition of sending command signal that represent a user intent to the user terminal.” (EX-1007, ¶82.) In addition, Barbara discloses that “[w]hen each utterance is received at the server, it is evaluated into a set of possible requests that the user may have actually meant.” (EX-1007, ¶90.)

In determining the interpretation of the received utterance, Barbara takes into account the utterance context. For example, Barbara discloses that “[w]hen each utterance is received at the server, it is evaluated into a set of possible requests that the user may have actually meant. A combined likelihood for each possible request is heuristically calculated. In the current implementation, this is based on a combination of:

the likelihood from the voice recognition engine that the text was actually what was said

the likelihood from the service's knowledge base that the request was what was meant given what was said

an estimation from the user's information base of which intents are likely.”

(EX-1007, ¶¶90-93.)

In Barbara, “[a]mbiguities are resolved when the system evaluates the likelihood that the user really is asking each of those things. For instance, 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. Contextual information such as this is captured as the user interacts with the system, in much the same way as interactions work in real life.” (EX-1007, ¶98.) Barbara explains that, for example, “[i]f 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.” (EX-1007, ¶98.) As another example, “[a] user who asks about their dog ten times a day might be considered very likely to be referring to it. Someone who is not involved in financial markets is unlikely to be asking about the Nikeii.” (EX-1007, ¶98.) That is, as Barbara explains, “[t]he history of what people have asked is heuristically used to assist in scoring the possible intents.” (EX-1007, ¶98.)

Specifically, Barbara discloses that “[t]he **information database may be used to help in determining which of the interpretations is correct.** For instance, when asked to display something, the correct query is much more likely to have some values than to have none. 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 or Tokyo the city, which the VR engine may have suggested as a possible hearing. The database has flea vaccination information for Toto but not for the others, so the score for the canine interpretation may be raised.

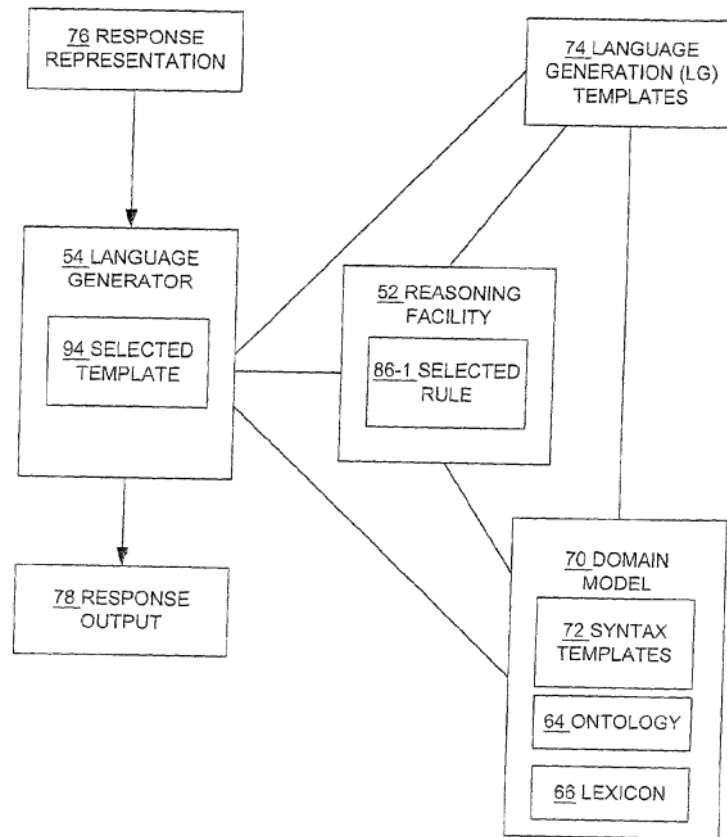
In this way the system ‘knows’ about the real world and can resolve ambiguities in many difficult sentences.” (EX-1007, ¶148.) Accordingly, “*an interpretation of the first natural language utterance*,” e.g., the word Toto refers to a dog, is “*determined based on the first context*,” e.g., “pet.”

## **8. Generating A Response [1G]/[10H]**

Barbara discloses “*generating, by the computer system, a first response to the first natural language utterance based on the interpretation*” [1G]/[10H].

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.) As discussed in §V.A.3, Barbara lacks details regarding system component used to generate a response.

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 combination discloses limitations [1G]/[10H].

### **C. Dependent Claims**

#### **1. Claims 2, 11**

Barbara discloses “*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.*”

Barbara 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.” (EX-1007, ¶168; *see also id.*, ¶¶86, 166-67.) 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 between a user and the computer system, 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.*”

Additionally, in the Barbara-Ross system, during a predetermined time, “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.”

## **2. Claims 3, 12**

Barbara discloses “*wherein the long-term knowledge comprises one or more of user preferences, ....*”

Barbara records “**personal details or preferences from what has been said**” in the “information database(s).” (EX-1007, ¶88.) For example, Barbara’s system infers that a speaker “who asks about their dog ten times a day” is a dog owner (e.g., shows a preference towards dogs as pets) 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.)

### 3. Claims 4, 13

#### a) Receiving a Second Natural Language Utterance [4A]/[13A]

Barbara discloses “receiving, by the computer system, a second natural language utterance in response to the first response” [4B]/[13B]. (EX-1003, ¶¶227-229.)

Barbara discloses that its system is a conversational system. For example, Barbara discloses that “[c]ontextual information such as this is captured as the user interacts with the system, in much the same way as interactions work in real life. 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.”** (EX-1007, ¶98.)

Accordingly, Barbara discloses that the system can receive a clarification (“*a second natural language utterance*”) in response to the system’s incorrect response (“*first response*”).

Moreover, Barbara notes that “[e]xpert systems attempt to distil the knowledge of an expert, such as a doctor or car mechanic, and produce a (large) set of rules so that the computer can do the job of the expert.” (EX-1007, ¶166-67.) As Barbara explains, “[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.) That is, Barbara discloses that the user can continually converse with the expert system.

#### **b) Identifying User Preferences**

Barbara discloses “*identifying [by the computer system] one or more user preferences based on the second natural language utterance.*” [4B]/[13B]. (EX-1003, ¶¶230-231.)

Barbara discloses that “[c]ontextual information ...is captured as the user interacts with the system .... 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.**”

(EX-1007, ¶98.) In this particular example, the system may capture that the user has a preference for dogs.

Barbara confirms that “[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.)

**c) Updating Long-Term Knowledge**

Barbara discloses “*updating, by the computer system, the long-term knowledge based on the identified user preferences*” [4C]/[13C]. (EX-1003, ¶232.)

As explained in §V.B.2, in the Barbara-Ross system, utterances of the current conversation and the result of their semantic analysis are stored in a conversational record, which represents short-term knowledge. However, as further explained in §V.B.4, after, a predetermined time when the information is purged from the “conversational record,” a POSITA would have understood that, at least part of it, e.g., “personal details or preferences from what has been said” (EX-1007, ¶88) would be moved into Barbara’s informational database, which represents long-term knowledge, because it constitutes “[t]he history of what people have asked [that is] used to assist in scoring the possible intents” (EX-1007, ¶98) and because it can be used to calculate Barbara’s combined likelihood (EX-1007, ¶90).

#### 4. Claims 5, 14

Barbara renders obvious “*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 first context for the first natural language utterance is further determined based on the profile associated with the first user.*” (EX-1003, ¶¶233-234.)

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, Barbara discloses “[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.** For example, software may be provided **to allow users to verbally enter personal details, such as the names of their children or their favorite sport or their mother's telephone number etc.** Non-personal information (such as weather reports or stock prices) may be directly fed into the information database by system operators.” (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. For example, as discussed in §V.B.4, 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” and “[a]n advantage of having the information database at the user's terminal is that the privacy of the user can be maintained” (EX-1007, ¶89), further confirming that “*the long-term knowledge is associated with a first user.*”

Barbara also renders obvious that “*the first context for the first 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,” and therefore “it is much more likely that they are referring to Toto the dog, rather than Toto the restaurant or Tokyo the city, which the VR engine may have suggested as a possible hearing.” (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. Accordingly, Barbara discloses that the “*first context for the first 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, 15**

Barbara discloses “*generating [by the computer system] one or more of a user profile... wherein subsequent natural language utterances are processed based at least on the one or more of the user profile....*”

As explained in §V.C.4, Barbara renders obvious “*generating a [user] profile associated with the first user based on the long-term knowledge.*” For the same reasons, Barbara also renders obvious “*generating one or more of a user profile.*”

Barbara further renders obvious that “*subsequent natural language utterances,*” e.g., the user asking for the date of Toto's last flea vaccination, “*are processed based at least on the user profile,*” e.g., information about Toto’s flea vaccination.

## **6. Claims 7, 16**

Barbara discloses “*receiving, by the computer system, information from a remote data source, wherein the long-term knowledge is based further on the information obtained from the remote data source.*”

The system of Barbara's Figure 6 (below) includes a main server (shaded pink) with "one or more service processors having software for receiving, interpreting and correcting the incoming information from the user terminal" (shaded yellow). (EX-1007, ¶85.) Barbara further discloses that "[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**. For example, software may be provided to allow users to verbally enter personal details, such as the names of their children or their favorite sport or their mother's telephone number etc." (EX-1007, ¶88.)

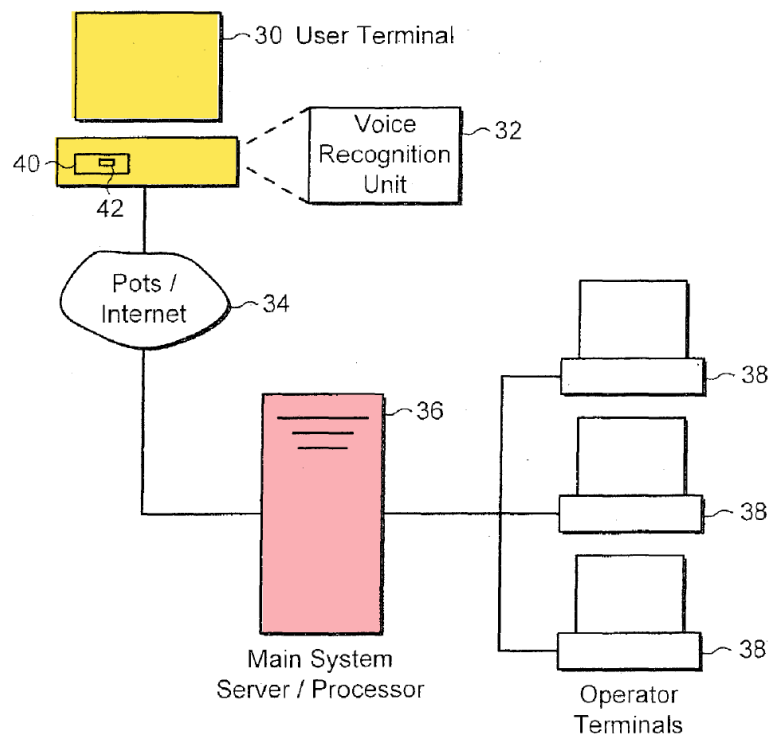


FIG. 6

Barbara, Figure 6

Accordingly, Barbara alone discloses that “*receiving by the computer system*” (server 36) “*information from a remote data source*” (user terminal 30) and that “*the long-term knowledge*” (information within informational database) “*is based further on the information obtained from the remote data source*” (personal details or preferences from information directly entered into the system by the user).

## **7. Claims 8, 17**

Barbara discloses “*determining, by the computer system, an interpretation of one or more recognized words of the first natural language utterance based on the first context.*”

As explained in §V.B.6, Barbara discloses an example where a “user asks for the date of Toto’s last flea vaccination.” (EX-1007, ¶148.) Barbara also discloses that “[c]ontextual information ... is captured as the user interacts with the system ....” (EX-1007, ¶98.) Barbara explains that when the user asks for Toto’s last flea vaccination, “it is much more likely that they are referring to Toto the dog, rather than Toto the restaurant or Tokyo the city, which the VR engine may have suggested as a possible hearing.” (EX-1007, ¶148.) That is, Barbara discloses “*determining an interpretation of one word*” (“Toto”) “*of the utterance, based on the first context,*” e.g., “pet” or “canine” context domain.

## 8. Claims 9, 18

Barbara discloses “*causing, by the computer system, the response to the first natural language utterance to be provided to a 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).)

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; *see also* EX-1007, ¶166.)

## **VI. Stipulation**

To simplify the *Fintiv* analysis, and following the precedential decision in *Sotera Wireless, Inc. v. Masimo Corp.*, Petitioner stipulates that, if IPR is instituted, Petitioner will not pursue in the related district court proceeding any ground that Petitioner raised or reasonably could have raised against the challenged claims during the instituted IPR.

## **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.*, 2:24-cv-00828-JRG-RSP (E.D. Tex., filed Oct. 9, 2024)
- *VB Assets, LLC v. Apple, Inc.*, No. 1:24-cv-01368-MN (D. Del., filed Dec. 13, 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)

**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:

| <b>Lead Counsel</b>                                                                                                                                                             | <b>Back Up Counsel</b>                                                                                                                                                                                       |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Lori Gordon (Reg. No. 50,633)<br>GOODWIN PROCTER LLP<br>1900 N St. N.W.<br>Washington, D.C. 20036<br>Phone: (202) 346-4000<br>Fax: (202) 346-4444<br>Gordon-ptab@goodwinlaw.com | Douglas Kline (Reg No. 35,574)<br>GOODWIN PROCTER LLP<br>100 Northern Avenue<br>Boston, MA 02210<br>Tel: (617) 570 1000<br>Fax: (617) 523-1231<br>dkline@goodwinlaw.com                                      |
|                                                                                                                                                                                 | Srikanth K. Reddy<br>( <i>pro hac vice</i> application to be filed)<br>GOODWIN PROCTER LLP<br>100 Northern Avenue<br>Boston, MA 02210<br>Tel: (617) 570 1000<br>Fax: (617) 523-1231<br>sreddy@goodwinlaw.com |
|                                                                                                                                                                                 | Brian T. Drummond (Reg. No. 68,414)<br>GOODWIN PROCTER LLP<br>100 Northern Avenue<br>Boston, MA 02210<br>Tel: (617) 570 1000<br>Fax: (617) 523-1231<br>bdrummond@goodwinlaw.com                              |
|                                                                                                                                                                                 | Theodoros Konstantakopoulos (Reg.<br>No. 74,155)<br>GOODWIN PROCTER LLP<br>620 Eighth Avenue<br>New York, NY 10018<br>Phone: (212) 813-8800<br>Fax: (212) 355-3333<br>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 1, 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. 10,510,341 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 12,449. 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 1, 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. 10,510,341 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 1, 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 facilitating natural language system responses utilizing accumulated short-term and long-term knowledge, the method being implemented by a computer system that includes

**[1P.2]** one or more physical processors executing one or more computer program instructions which, when executed, perform the method, the method comprising:

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

**[1B]** expiring, by the computer system, one or more items of short-term knowledge that are based on one or more natural language utterances received prior to the predetermined time period;

**[1C.1]** accumulating, by the computer system, long-term knowledge based on one or more natural language utterances received prior to the predetermined time period,

**[1C.2]** wherein the long-term knowledge includes at least one of the one or more expired items of short-term knowledge;

**[1D]** receiving, at the computer system, a first natural language utterance via an input device;

**[1E]** determining, by the computer system, based on the short-term knowledge and the long-term knowledge, a first context for the first natural language utterance;

**[1F]** determining, by the computer system, based on the first context, an interpretation of the first natural language utterance; and

**[1G]** generating, by the computer system, a first response to the first natural language utterance based on the interpretation.

**[2]** The method of claim 1, 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.

**[3]** The method of claim 1, wherein the long-term knowledge comprises one or more of user preferences, an indication of utilized domain agents, contexts of prior natural language utterances, requests of prior natural language utterances, and tasks performed in response to prior natural language utterances.

**[4A]** The method of claim 1, the method further comprising: receiving, by the computer system, a second natural language utterance in response to the first response;

**[4B]** identifying, by the computer system, one or more user preferences based on the second natural language utterance; and

**[4C]** updating, by the computer system, the long-term knowledge based on the identified user preferences.

**[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 first context for the first natural language utterance is further determined based on the profile associated with the first user.

**[6]** The method of claim 1, the method further comprising: generating, by the computer system, one or more of a user profile, an environmental profile, a historical profile, and/or a cognitive profile based on the long-term knowledge, wherein subsequent natural language utterances are processed based at least on the one or more of the user profile, the environmental profile, the historical profile, and the cognitive profile.

**[7]** The method of claim 1, the method further comprising: receiving, by the computer system, information from a remote data source, wherein the long-term knowledge is based further on the information obtained from the remote data source.

**[8]** The method of claim 1, wherein determining the interpretation of the first natural language utterance comprises: determining, by the computer system, an interpretation of one or more recognized words of the first natural language utterance based on the first context.

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

**[10P]** A system for facilitating natural language system responses utilizing accumulated short-term and long-term knowledge, the system comprising:

**[10A]** one or more physical processors programmed with one or more computer program instructions which, when executed, configure the one or more physical processors to:

**[10B]** accumulate short-term knowledge based on one or more natural language utterances received during a predetermined time period;

**[10C]** expire one or more items of short-term knowledge that are based on one or more natural language utterances received prior to the predetermined time period;

**[10D.1]** accumulate long-term knowledge based on one or more natural language utterances received prior to the predetermined time period,

**[10D.2]** wherein the long-term knowledge includes at least one of the one or more expired items of short-term knowledge;

**[10E]** receive a first natural language utterance via an input device;

**[10F]** determine a first context for the first natural language utterance based on the short-term knowledge and the long-term knowledge;

**[10G]** determine an interpretation of the first natural language utterance based on the first context; and

**[10H]** generate a first response to the first natural language utterance based on the interpretation.

**[11]** The system of claim 10, 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.

**[12]** The system of claim 10, wherein the long-term knowledge comprises one or more of user preferences, an indication of utilized domain agents, contexts of prior natural language utterances, requests of prior natural language utterances, and tasks performed in response to prior natural language utterances.

**[13A]** The system of claim 10, wherein the one or more physical processors are further configured to: receive a second natural language utterance in response to the first response;

**[13B]** identify one or more user preferences based on the second natural language utterance; and

**[13C]** update the long-term knowledge based on the identified user preferences.

**[14]** The system of claim 10, 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 first context for the first natural language utterance is further determined based on the profile associated with the first user.

**[15]** The system of claim 10, wherein the one or more physical processors are further configured to: generate one or more of a user profile, an environmental profile, a historical profile, and/or a cognitive profile based on the long-term knowledge, wherein subsequent natural language utterances are processed based at

least on the one or more of the user profile, the environmental profile, the historical profile, and the cognitive profile.

[16] The system of claim 10, wherein the one or more physical processors are further configured to: receive information from a remote data source, wherein the long-term knowledge is based further on the information obtained from the remote data source.

[17] The system of claim 10, wherein to determine the interpretation of the first natural language utterance, the one or more physical processors are configured to: determine an interpretation of one or more recognized words of the first natural language utterance based on the first context.

[18] The system of claim 10, wherein the one or more physical processors are further configured to: cause the response to the first natural language utterance to be provided to a user.