

**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

---

IPR2025-00868

U.S. Patent No. 8,515,765

---

**DECLARATION OF STUART LIPOFF IN SUPPORT OF PETITION FOR  
*INTER PARTES* REVIEW OF U.S. PATENT NO. 8,515,765**

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

|                                                                                                            | Page |
|------------------------------------------------------------------------------------------------------------|------|
| I. QUALIFICATIONS .....                                                                                    | 2    |
| II. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES .....                                                       | 8    |
| III. '765 PATENT .....                                                                                     | 11   |
| A. Technical Background .....                                                                              | 11   |
| 1. Introduction to Spoken Dialog Systems .....                                                             | 11   |
| 2. Basic Architecture of SDSs .....                                                                        | 13   |
| 3. Dialog Management .....                                                                                 | 16   |
| 4. Initiative Management .....                                                                             | 18   |
| 5. Control of Dialog Flow .....                                                                            | 19   |
| 6. Classification of Different Spoken Dialogue Systems .....                                               | 23   |
| B. Overview of the '765 Patent .....                                                                       | 30   |
| C. Disclosures Relevant to the Claims of the '765 Patent .....                                             | 34   |
| D. Prosecution History .....                                                                               | 37   |
| E. Level of Ordinary Skill in the Art .....                                                                | 40   |
| F. Claim Construction .....                                                                                | 41   |
| IV. GROUND 1: COMBINATION OF SMITH & HIPPI AND GUINN<br>RENDERS OBVIOUS CLAIMS 1-14 .....                  | 42   |
| A. Overview of the Combination .....                                                                       | 42   |
| 1. Overview of Smith & Hipp .....                                                                          | 42   |
| 2. Overview of Guinn .....                                                                                 | 49   |
| 3. Motivation to Combine .....                                                                             | 52   |
| B. Independent Claims 1 and 10 .....                                                                       | 56   |
| 1. Preamble [1P] / Preamble [10P] .....                                                                    | 58   |
| 2. Receiving Natural Language Utterance [1A] / [10A] .....                                                 | 62   |
| 3. Determining the Conversation Type Corresponding to the<br>Natural Language Utterance [1B] / [10B] ..... | 67   |
| 4. Generating a Response to the Natural Language<br>Utterance [1C] / [10C] .....                           | 76   |

## TABLE OF CONTENTS

(continued)

Page ii

|    |                                                                                                                                                                        |     |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| C. | Claims 2 and 11 – “format of the response is adapted to provide discrete piece of information, ... a plurality of pieces of information, or obtain information.” ..... | 79  |
| D. | Claim 3 – “discrete” responses shorter than “plurality of pieces” responses .....                                                                                      | 94  |
| E. | Claims 4 and 12 – “determining conversation type” based on “conversational goal.” .....                                                                                | 97  |
| F. | Claims 5 and 13 – “conversation type” determined based on “role” .....                                                                                                 | 102 |
| G. | Claims 6, 7, and 14 – “conversation type” determined based on “allocation of information.” .....                                                                       | 107 |
| H. | Claim 8 - “only provides information requested ... limits the user’s future input” .....                                                                               | 116 |
| I. | Claim 9 - “user controls the conversation or ... provides information” .....                                                                                           | 119 |
| V. | CONCLUSION.....                                                                                                                                                        | 123 |

I, Stuart Lipoff, declare as follows:

1. I have been retained by Goodwin Procter LLP on behalf of Petitioners Samsung Electronics Co. Ltd. and Samsung Electronics America, Inc. (“Petitioners”) to provide this Declaration concerning technical subject matter relevant to the petition for *Inter Partes* Review (“Petition”) of U.S. Patent 8,515,765 (“the ’765 patent”).

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

3. I have been asked to provide my technical opinions regarding how a person of ordinary skill in the art would have understood the claims of the ’765 patent at the time of the alleged invention, which I have been asked to assume is the 2006 timeframe. For purposes of whether the teachings of the prior art render the claims of the ’765 patent obvious, I have been asked to assume the date of October 16, 2006. I have also been asked to provide my technical opinions on how concepts in the ’765 patent specification relate to claim limitations of the ’765 patent.

4. In reaching the opinions provided herein, I have considered the ’765 patent, its prosecution history, and the references cited in my Declaration and the Appendix. I have also drawn on my own education, training, research, knowledge, and personal and professional experience.

5. In general, I have been asked to cite to the specification of a patent or patent publication using the following formats: (Patent, Col:Line Number(s)) or (Patent, Paragraph Number(s)). For example, the citation ('765 patent, 1:1-10) points to the '765 patent specification at column 1, lines 1-10. Also, for convenience, I have been asked to use italics to denote limitations from the challenged claims. Unless otherwise noted, all emphasis is added.

6. I have personal knowledge of the facts and opinions set forth in this declaration and believe them to be true. If called upon to do so, I would testify competently thereto. I have been warned that willful false statements and the like are punishable by fine or imprisonment, or both.

7. I am being compensated for my time at my standard consulting rate. I am also being reimbursed for expenses that I incur during the course of this work. My compensation is not contingent upon the results of my study and analysis, the substance of my opinions, or the outcome of any proceeding involving the '765 patent. I have no financial interest in the outcome of this matter or in any litigation involving the '765 patent.

## **I. Qualifications**

8. I believe I am well qualified to render useful opinions on this matter. I will briefly summarize my knowledge, training, and experience here. A more

detailed summary of my background, education, experience, and publications is set forth in my curriculum vitae (CV), which is submitted as EX-1004.

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

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

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

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

13. I am a fellow of the Institute of Electrical and Electronics Engineers (“IEEE”) Consumer Electronics, Communications, Computer, Circuits, and Vehicular Technology Groups. I have been a member of the IEEE Consumer Electronics Society National Board of Governors (formerly known as the

Administrative Committee) since 1981, and I was Boston Chapter Chairman of the IEEE Vehicular Technology Society from 1974 to 1976. I served as the 1996-1997 President of the IEEE Consumer Electronics Society, and from 1999 to 2018 I served as Chairman of the Society's Technical Activities and Standards Committee and as Vice President of Publications for the Society. From 2018 to 2023 I served as Vice President of Standards and Industry Activities for the Society and currently serve on the Board of Governors as the Historian for the Society. I have also served as a Mazura Ibuka Customer Technology Award committee member for the IEEE's Award in the field of consumer electronics.

14. I have prepared and presented numerous papers at the IEEE and at other professional meetings. For example, in fall 2000, I served as general program chair for IEEE's Vehicular Technology Conference on advanced wireless communication technology. I have organized sessions at The International Conference on Consumer Electronics, and I was the 1984 program chairman. I conducted an eight-week IEEE-sponsored short course on Fiber Optics System Design. I received IEEE's Centennial Medal in 1984, and I received IEEE's Millennium Medal in 2000.

15. As Vice President and Standards Group Chairman for the Association of Computer Users ("ACU") from 1980 to 1983, I served as the ACU representative to the ANSI X3 Standards Group. From 1976 to 1978, I served as Chairman of the

task group on user rule compliance for the FCC's Citizens Advisory Committee on Citizen's Band radio.

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

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

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

interactions. I also have extensive experience in foundation technologies and the applications supporting speech signal processing.

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

20. For approximately four years, from 1972 to 1976, I served as Section Manager for Bell & Howell Communications Company, where I also had project design responsibilities for paging and wireless communication products. I supported projects developing covert audio intelligence systems that recognized speech and activated digital voice compression recording systems. I also led projects for voice-based radio paging systems that recorded speech input, processed the speech to remove silence, processed the speech to digitally compress the speech, and store and forward the speech upon demand from dual tone multi-frequency or computer keyboard retrieval from the servers.

21. For 25 years from 1976 to 2001, I worked for Arthur D. Little, Inc. ("ADL"), where I became the Vice President and Director of Communications,

Information Technology, and Electronics (CIE) and served in that role for 10 years, from 1991 to 2001. At ADL, I was responsible for the firm's global CIE practice in laboratory-based contract engineering, product development, and technology-based consulting. I was also involved in multiple pioneering efforts to identify and explore customer-to-business and business-to-business electronic commerce and transactions information processing opportunities, or "e-commerce". These projects involved technology assessment and analysis as well as developing architectures and systems to support multiple applications. The technologies typically involved an information retrieval component.

22. While at ADL, I worked on several projects involving the combination of voice interfaces (including speech recognition and voice audio output) and information retrieval. Over the course of three years in the early-1990s, I worked on a project for Bolt Beranek and Newman (BB&N), where I evaluated and benchmarked technology for a voice input/output application that allowed end users (*e.g.*, travel agents) to use speech inputs to interact with airline reservation databases to retrieve information about travel reservation options, where the results were returned to the user in an audible message. This system included a natural language front-end speech-interface module with speech recognition that used pre-defined recognition grammars to convert the end user's speech into structured commands supported by an airline reservation system. As another example, over the course of

three years in the mid-1990s, I also worked on a project for Texas Instruments that applied a speech–recognition interface for a variety of applications that retrieved information from database servers.

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

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

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

26. Over the course of 15 years starting in the early 1980s, I worked on a project for the USPS, where we developed a real-time automated postal teller system that served as an interface between end-users and the USPS’s information systems. This system included voice prompts for the vision impaired.

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

27. I am not a lawyer, and I will not provide any legal opinions. Although I am not a lawyer, I have been advised certain legal standards are to be applied by

technical experts in forming opinions regarding the meaning and validity of patent claims.

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

29. I have been informed that a patent claim is invalid if it would have been obvious to a person of ordinary skill in the art. In analyzing the obviousness of a claim, I understand the following factors may be taken into account: (1) the scope and content of the prior art; (2) the differences between the prior art and the claims; (3) the level of ordinary skill in the art; and (4) any so called “secondary considerations” of non-obviousness, if they are present. I am not aware of any evidence of secondary considerations of non-obviousness relevant to the ’765 patent. I reserve the right to supplement this Declaration if Patent Owner (“PO”) introduces evidence of secondary considerations of non-obviousness.

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

(1) identify the particular references that, singly or in combination, make the patent obvious;

(2) specifically identify which elements of the patent claim appear in each of the asserted references; and

(3) explain why a person of ordinary skill in the art would have combined the references, and how they would have done so, to create the inventions claimed in the patent. I further understand that exemplary rationales that may support a conclusion of obviousness include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- use of known technique(s) to improve similar devices (methods or products) in the same way;
- applying a known technique to a known device (method or product) ready for improvement to yield predictable results;
- “obvious to try” – choosing from a finite number of identified, predictable solutions with a reasonable expectation of success; known work in one field of endeavor may prompt variations of the work for use in either the same field or a different field based on design incentives or other market forces if the variations are predictable to a person of ordinary skill in the art; and
- some teaching, suggestion, or motivation in the prior art that would have led a person of ordinary skill in the art to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

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

### **III. '765 Patent**

32. The '765 Patent, filed on October 3, 2011, “relates to a cooperative conversational model for a human to machine voice user interface.” (EX-1001, 1:17-18.) The '765 patent claims priority to non-provisional U.S. Application No. 11/580,926, filed on October 16, 2006.

33. By 2006, cooperative conversation systems, referred to as “spoken dialog systems” or “interactive dialog systems”, were well-known and in use, as I discuss further in §III.A below. Indeed, as I also discuss below, the '765 patent and the challenged claims do nothing more than co-opt well-known concepts in the art.

#### **A. Technical Background**

34. Prior to the earliest priority date of the '765 patent (October 16, 2006), conversational systems like the type described in the '765 patent had been known and in use for at least a decade. In the sections below I introduce the reader to the state of the relevant art as of 2006.

##### **1. Introduction to Spoken Dialog Systems**

35. Spoken Dialog Systems (SDSs), sometimes called interactive speech systems, are “computer systems with which humans interact on a turn-by-turn basis.” (Sebastian Möller, *Quality of Telephone-Based Spoken Dialogue Systems* (2005) (“Möller”); EX-1008, 2.) At a high level, the task of an interactive dialog

system “is to enable and support the spoken interaction between the human user and the service offered by the application system.” (EX-1008, 19.) “They enable and support the communication of information between two parties (agents), mostly between a human user and a machine agent,” sometimes referred to as a system or computer. (EX-1008, 2.) “[T]he aim is to model naturally occurring conversation.” (Michael F. McTear, *Spoken Dialogue Technology* (2004) (“McTear”); EX-1009 45.)

36. Some of the earliest spoken dialogue systems were in the realm of science fiction. The ship’s computer in Star Trek and HAL in 2001: A Space Odyssey are two that most prominently captured the attention of popular culture as the epitome of “the future.” Indeed, Alan Turing famously proposed in 1950 that the ultimate hallmark of computer intelligence would be its ability to sustain a conversation indistinguishable from that of a human, now known as the Turing Test. (Xuedong Huang et al., *Spoken Language Processing* (2001); EX-1010, 3.) As a result of decades of investment into research of these systems from industry, academia, and government, even years before the priority date of the ’765 patent, that day was “verging on reality.” (EX-1009, 16.)

37. Namely, “[s]ince the early 1970s, researchers at AT&T, BBN, CMU, IBM, Lincoln Labs, MIT, and SRI have made major contributions in Spoken Language Understanding Research.” (EX-1010, xxi.) In 1971, the Defense

Advanced Research Projects Agency (DARPA) initiated a successful five-year effort spanning institutions across the country to develop speech understanding systems. (EX-1010, xxi.)

38. By the early 2000s, an entire discipline of “spoken dialogue technology” had emerged. (EX-1009, vii.) This was also fueled by “an increasing commercial interest in the technology, with a number of large, influential companies actively involved in spoken language technology and offering advanced toolkits and development platforms to researchers wishing to explore the exciting potential of conversing with a computer using spoken natural language.” (EX-1009, vii.) DARPA continued to fund influential projects like the MIT Galaxy and the CU Communicator projects, which emphasized dialogue management and user modeling. (EX-1009, 324-32.) Meanwhile, Möller highlights that by 2003, “about one third of the automatized telephone lines” were expected to include speech recognition capabilities, reflecting a growing reliance on SDSs for customer service and call routing. (EX-1008, 1.)

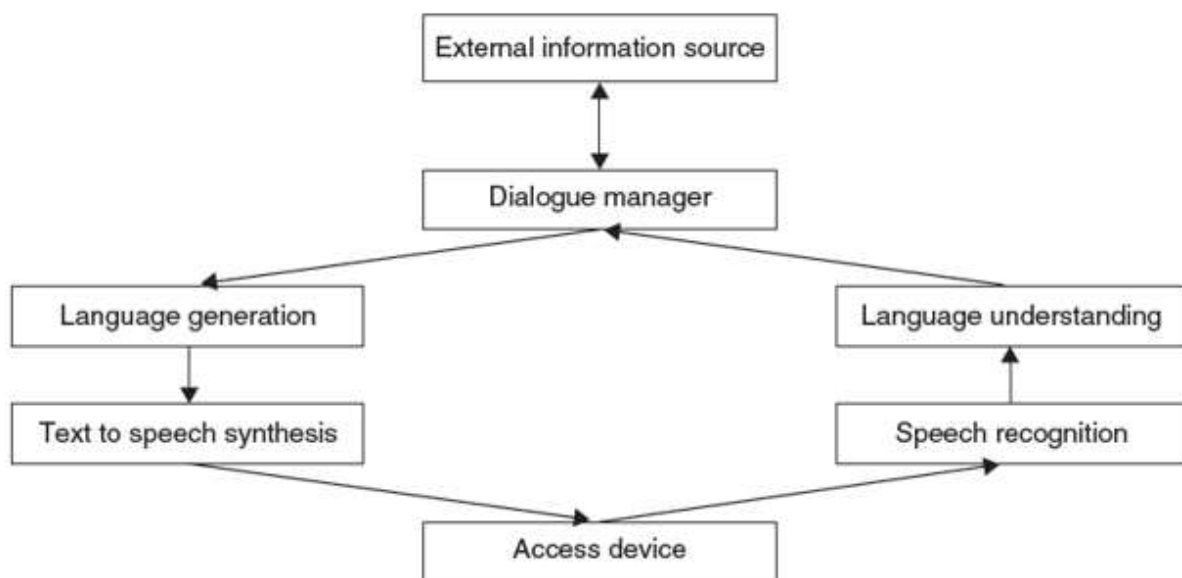
## **2. Basic Architecture of SDSs**

39. All SDSs are made up of certain common “major components.” (*See, e.g.,* EX-1008, 20.) These include:

- **Speech Recognition:** Converts the spoken input into a word string or word hypothesis graph.

- **Language Understanding:** Interprets the recognized word string semantically, producing a semantic frame.
- **Dialogue Manager:** Interprets the semantic frame in context, manages dialogue history, and executes dialogue control.
- **Language Generation:** creates natural language replies based on the dialogue manager's output.
- **Text-To-Speech Synthesizer:** Converts the response into speech, and transmits it back to the user.

40. The components of an SDS “may be implemented in different ways.” (EX-1008, 21.) McTear provides one common framework, which is represented below:

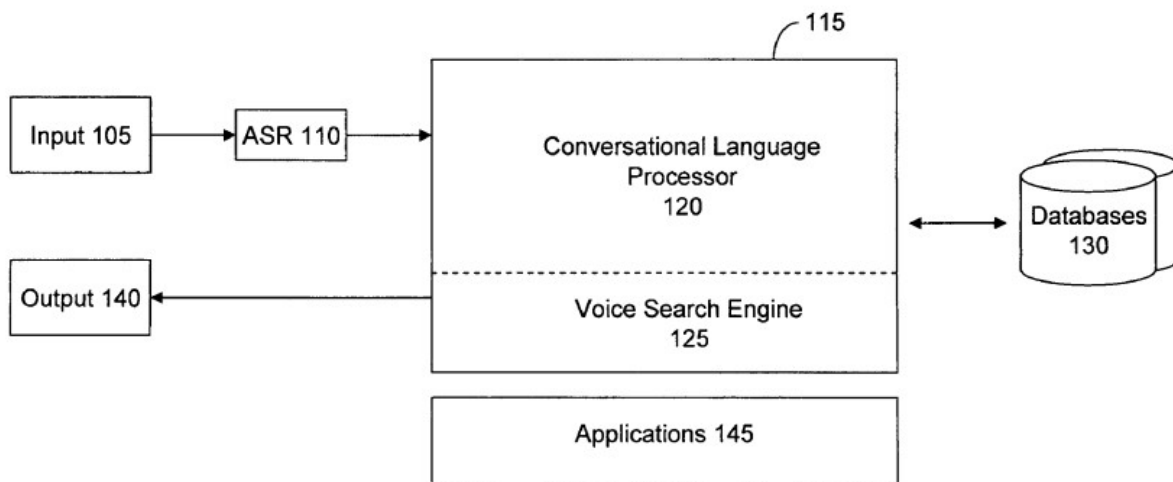


**McTear, Figure 4.1 (80)**

41. These components are also often paired with other components as well (e.g., a module allowing the system to interface with a phone line to support

telephony-based operations or a dialogue history to allow for the system to draw from previous user interactions to facilitate operation of the SDS. (EX-1008, Fig. 2.4 (20).)

42. The components can be organized based on the needs of the application. For example, the components can be divided “into a speech platform (containing the phone server, the speech recognizer and the speech synthesizer), a voice application server (similar to a web application server, and containing the semantic analysis, the dialogue management and the response generation), and the application back-end (e.g., the database).” (EX-1008, 21.) I note this implementation is similar to the implementation described in the ’765 patent that I discuss in §III.B and is illustrated in Figure 1 of the ’765 patent (reproduced below).



**'765 Patent, Figure 1**

### 3. Dialog Management

43. In a spoken dialog system, the dialog manager is the “central component.” (EX-1009, 107.) The dialog manager “accepts spoken input from the user, produces messages to be communicated to the user, interacts with external knowledge sources, and generally controls the dialogue flow.” (EX-1009, 107.) Put differently, according to Möller, “[i]t is the task of the dialogue manager to guarantee the smooth course of the dialogue, so that it is coherent with the task, the domain, the history of the interaction, with general knowledge of the ‘world’ and of conversational competence, and with the user.” (EX-1008, 27.)

44. Dialogue control presents a significant challenge in human-machine interaction because human dialogue is inherently flexible, context-dependent, and often ambiguous—features that are difficult to replicate or interpret computationally. (See EX-1008, 3.) As Möller explains, “the principles underlying this type of interaction are different from the ones which govern [] conversations between humans, because of the limitations of the machine interaction partner.” (EX-1008, ix.) Put differently, humans bring to every conversation a wealth of contextual knowledge, flexible reasoning, and conversational norms that machines simply do not possess.

45. Even before the earliest priority date of the ’765 patent, it was known that providing a useful and seamless experience despite the intrinsic limitations of

computer systems was the job of the dialog manager. The dialog manager must strike a balance: it must allow enough flexibility to accommodate natural, spontaneous speech, while also enforcing enough structure to accommodate the limitations of the system. (See EX-1009 111-16.)

46. In furtherance of its role, the dialog manager in speech dialog systems takes on numerous responsibilities. Core functions include:

- the collection of all information from the user which is needed for the task,
- the distribution of dialogue initiative,
- the provision of feedback and verification of information understood by the system,
- the provision of help to the user,
- the correction of errors and misunderstandings,
- the interpretation of complex discourse phenomena like ellipses and anaphoric references, and
- the organization of information output to the user.

(EX-1008, 27.)

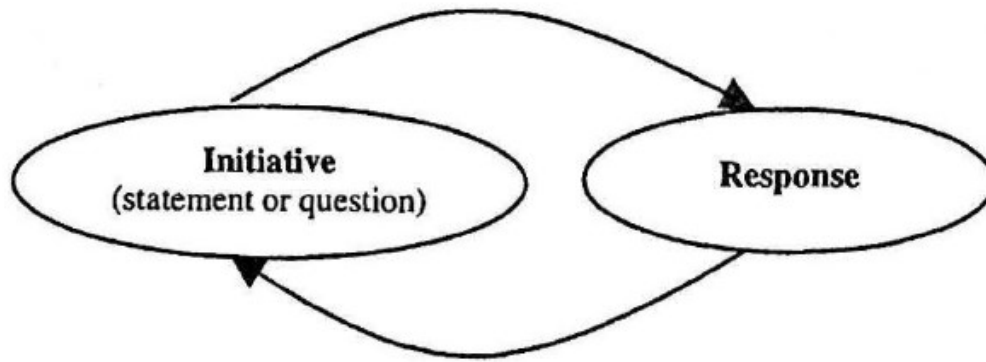
47. At bottom, though, the dialog manager's primary role is to mediate between user input, system capabilities, and the application logic, enabling the system to conduct multi-turn conversations while managing ambiguity and error correction. As Huang put it, "the dialogue manager controls the flow of conversation" and is "the central component that communicates with applications

and the spoken language understanding modules such as discourse analysis, sentence interpretation, and response generation.” (EX-1010, 7.)

48. In the next section I discuss two related aspects of the dialog manager that are particularly relevant to the issues concerning the ’765 patent: Initiative Management and Dialog Flow.

#### **4. Initiative Management**

49. In an interactive system, the “dialogue may be system-led, user-led or mixed initiative.” (EX-1009, 107; *see also* EX-1010, 860.) “In a system-led dialogue the system asks a sequence of questions to elicit the required parameters of the task from the user.” (EX-1009, 107.) “In a user-led dialogue the user controls the dialogue and asks the system questions in order to obtain information.” (EX-1009, 107.) “In a mixed initiative dialogue control is shared” such that the “user can ask questions at any time, but the system can also take control to elicit required information or to clarify unclear information.” (EX-1009, 107.) Regardless of the initiative type, “the fundamental structure of dialog consists of initiative-response pairs,” shown in Huang’s Figure 17.2 (below). (EX-1010, 860.)



**Huang, Figure 17.2 (860)**

50. “These distinctions are reflected in the methods that can be used to control the dialogue flow.” (EX-1009, 107.) For example, in system-led implementations, control involves determining what and when questions should be asked by the system. (EX-1009, 107.) In system-led implementations, the dialog flow “can be scripted as a sequence of choices.” (EX-1009, 107.) However, in the mixed-initiative, more open-ended implementations, “the choice of the next action is determined dynamically, based on the current state of the dialogue.” (EX-1009, 107-108.) In such systems, the dialog manager “acts as an intelligent agent that makes rational decisions about how to manage the dialog.” (EX-1009, 107-108.)

## **5. Control of Dialog Flow**

51. In spoken dialogue systems, the control of dialogue flow is a core responsibility of the dialog manager. Though they go by different names, there are largely three ways control of flow can be achieved. McTear refers to finite-state-based, frame-based, and agent-based dialogue controls. (EX-1009, 111.) Möller, in contrast, refers to dialogue grammar, plan-based approaches, and collaborative

approaches. (EX-1008, 28.) Though there are differences between these two, I will focus on describing the approaches using Möller’s framework.

52. Starting with dialogue grammars, this model represents the most structured and rigid form of dialogue flow control, which implement dialogue flow using structures like graphs or finite-state machines. As Möller explains, dialogue grammars implement a top-down model of dialogue management using structures such as graphs, finite-state machines, or declarative grammar rules. Each node in the graph corresponds to a specific system prompt—ranging from closed or open questions to explanations or choices delivered via “audible quoting”—and transitions are determined by the semantic interpretation of the user’s input, constrained by a context-free grammar. (EX-1008, 28.) This model supports simple, robust, and user-guided interactions, making it well-suited for well-structured tasks. However, Möller emphasizes its limitations: “a lack of flexibility, and a very close relation or mixture of task and dialogue models,” which renders it unsuitable for “ill-structured tasks” or “complex transactions.” (EX-1008, 28.) He notes that the rigid, system-driven structure of dialogue grammars can be mitigated through frame-based approaches, which offer greater adaptability by organizing application needs hierarchically. (EX-1008, 28.)

53. Next are plan-based approaches. These focus on modeling the communicative goals behind user interactions, including any relevant sub-goals.

These systems use plan operators to analyze dialogue structure and infer the user's underlying intentions, enabling them to interpret even indirect speech acts. While offering greater flexibility and depth of understanding than simpler models like dialogue grammars, plan-based systems are also significantly more complex. A key requirement is alignment between the plans of the human and system agents—if these diverge, the dialogue may “head in the completely wrong direction.” (EX-1008, 28.)

54. Finally, collaborative approaches to dialogue management move beyond task structure to focus on the motivation and mechanics of dialogue itself. Unlike plan-based models that center on predefined goals, collaborative systems aim to model the shared beliefs and mutual understanding between participants. The dialogue manager tracks what both the system and the user believe to be true within the conversation, treating “accepted goals [as] shared beliefs.” (EX-1008, 28.) These approaches typically blend elements from agent theory, plan-based reasoning, and dialogue grammars to model generic, flexible properties of dialogue. However, this openness comes at a cost: because the dialogue is less constrained, users are more likely to speak in unanticipated ways, increasing the complexity of interpretation. As a result, collaborative systems require more advanced natural language understanding and reasoning capabilities. (EX-1008, 29.)

55. McTear provides additional detail on this “collaborative approach” system, which he refers to as “agent-based” control implementations. (EX-1009, 116.) As McTear describes, these systems often “draw on techniques from Artificial Intelligence (AI) and focus on the modelling of dialogue as collaboration between intelligent agents to solve some problem or task.” (EX-1009, 116.) In these systems, communication “is viewed as interaction between two agents, each of which is capable of reasoning about its own actions and beliefs, and sometimes also about the actions and beliefs of the other agent.” (EX-1009, 117.) “The dialogue model takes the preceding context into account with the result that the dialogue evolves dynamically as a sequence of related steps that build on each other.” (EX-1009, 117.) These systems may also “use expectations to predict and interpret the user’s next utterances.” (EX-1009, 117.)

56. Generally, agent-based, or collaborative, systems “require complex dialogue and user models as well as mechanisms for using these models as a basis for decisions on how to control the dialogue.” (EX-1009, 124.) “Information about the dialogue history and the user can be used 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. These sorts of modelling involve representations of discourse structure, of intentions, goals and beliefs, and of dialogue as a collaborative activity.” (EX-1009, 124.)

57. In the next section I discuss three different configurations for speech dialogue systems and how they distinctly configure initiative and dialog flow to support each configuration's designs and needs.

## **6. Classification of Different Spoken Dialogue Systems**

58. There are many different ways to classify speech systems, reflecting the complexity and interdisciplinary nature of the field. As Bernsen notes, “there simply is no complete, general and accepted theory yet,” and instead we are left with “an organized conceptual toolbox of elements, at least some of which need to be taken into consideration when developing today's interactive speech systems.” (Niels Ole Bernsen et al., *Designing Interactive Speech Systems: From First Ideas to User Testing* (1998) (“Bernsen”); EX-1013, 31.) This diversity of elements and viewpoints means that classification schemes often vary depending on the intended use, the theoretical background, or the level of abstraction. As such, any model or taxonomy of speech systems must be viewed as one of many possible frameworks rather than a definitive structure.

59. According to Bernsen and McTear, one way to categorize different systems is by how they perform, which Bernsen calls “the performance layer.” As he describes it: the performance layer represents the “observable behavior of the system during the interaction.” (EX-1013, 55-56.) It captures how the system functions in real-time use and is defined by three key features: “cooperativity,

initiative, and the system’s influence on user behavior.” (EX-1013, 56.) Cooperativity is crucial for supporting effective, task-oriented dialogue; initiative refers to how control shifts between user and system—ranging from system-led to user-led to mixed; and influencing user behavior highlights how system design and output can shape user responses and interaction patterns.

60. While there are many ways to categorize spoken dialogue systems across the dimensions of cooperativity, initiative, and influencing user behavior, three primary categories of dialogue systems have emerged in the literature: information-seeking, instructional, and conversational. Each of these three categories reflects a distinct distribution of initiative between user and system, and affects the architecture and dialogue strategies accordingly.

**i. Information Retrieval / Task-Oriented Dialogues**

61. In information-seeking systems, the user takes the initiative. These systems are designed for users who already know what they want and can articulate their goal through direct queries. The dialogue system, therefore, must be responsive, precise, and efficient, often relying on tightly scoped input and robust response generation.

62. As Möller explains, “[m]ost of the currently available [spoken dialogue] systems enable a task-orientated dialogue, i.e. the goal of the interaction

is fixed to a specific task which can only be reached if both interaction partners cooperate.” (EX-1008, 3.)

63. McTear concurs, characterizing these systems as those where “the user asks one or more questions that the system interprets and answers.” (EX-1009, 109.) In this way, “a user-directed dialogue is rather like a natural language interface to a database.” (EX-1009, 110.) These systems were typically utilized within a constrained domain such as train schedules, stock prices, or directory assistance.

64. The dialogue systems developed under the Air Traffic Information System (ATIS) projects and the D’Homme project discussed in McTear provide an example of such a system. An example dialogue is below:

**1 User:** How many employees living in the London area earn more than £50,000?

**2 System:** Fifty four.

**3 User:** How many are female?

**4 System:** Eleven.

**5 User:** And managers?

**6 System:** Nine.

(EX-1009, 109.)

65. As McTear further explains, “[i]n a user-directed dialogue the user determines the questions to be asked and the role of the system is to attempt to answer the questions. In some applications, the system may ask clarification questions, if some aspect of the user’s question is unclear.” (EX-1009, 109.)

66. This category of spoken-dialogue systems covers many common telephony-based applications from the late 1990s and early 2000s. An archetype of this approach is the interactive voice response (IVR) system. IVRs are often used for queries such as checking account balances or booking tickets.

## **ii. Instructional or Tutorial Dialogues**

67. Instructional systems adopt a system-led dialogue model whose objective is to convey knowledge from the system to the user. Here, the system takes primary initiative and guides the user through a process, often in the context of learning, onboarding, or walkthroughs.

68. By 2007, an entire subfield related to instructional or tutorial speech dialogue systems had developed, called intelligent tutoring systems (“ITS”). These systems rely on instructional dialogue to incrementally guide students through conceptual learning. For instance, in the 2004 ITS Conference proceedings, systems like AutoTutor and RMT exemplify instructional dialogue patterns. In these, the system poses problems, scaffolds user understanding, and controls the dialogue structure to ensure learning objectives are met. (See Peter Wiemer-Hastings et al., *RMT: A Dialog-Based Research Methods Tutor With or Without a Head*, 7th International Conference on Intelligent Tutoring Systems, Vol. 3220 (Jan. 2004); EX-1011; see also Arthur C. Graesser et al., *AutoTutor: A Tutor with Dialogue in*

*Natural Language*, 36(2) Behavior Research Methods, Instruments, & Computers 180 (2004); EX-1012.)

69. Tutorial systems frequently employ spoken dialogue to simulate a teaching experience. For example, RMT, (research methods tutor), uses spoken natural language prompts to explain and clarify key ideas. (EX-1011, 1.) The dialogue follows a system-led script. The system's design is centered around short, structured turns where the tutor asks questions, gives hints, or provides brief explanations. Another such system is called the AutoTutor. AutoTutor presents challenging conceptual questions and controls the dialogue using a finite-state dialogue manager that maps learner input to pedagogical moves. Although the learner can respond in natural language, the system largely controls the progression of the conversation (EX-1012, 180-81.)

70. The importance of initiative and dialog control is emphasized in instructional / tutorial systems. As explained by Bernsen: "[Early] experiments were made with free mixed initiative dialogue openings of the 'Can I help you?' type. This opening turned out to strongly invite human-human style, lengthy and complex accounts from users which the system had no chance of understanding." (EX-1013, 57.)

71. Different tutorial systems took different strategies to address this problem. For example, RMT limits learner control by expecting narrowly defined

answers and using graph-based scripting to decide its next move. (EX-1011, 1-2.) Similarly, AutoTutor maintains control through a Socratic method that combines “prompts,” “pumps,” and “hints” to guide learners through reasoning tasks without allowing deviation from core content. (EX-1012, 180.) As described, AutoTutor presents challenging problems (formulated as questions) from a curriculum script and then engages in mixed initiative dialogue that guides the student in building an answer.” (EX-1012, 180.) The system often responds with moves like “What else?” (a pump), or “I think you can get this. [What about ...?]” (a prompt) to shape learner responses. (EX-1012, 183-84.)

72. The techniques employed by these systems to maintain dialogue control trace back to work done decades previously. For example, in the 1980s, Whittaker and Stenton described their observations as to the types of statements that were associated with obtaining or relinquishing control. As Bernsen later summarized their work: Whittaker and Stenton found that “the speaker has control in a request unless followed by a request or directive; in an assertion unless it is a response to a request; and in directives (commands); the listener has (gets) the control in a prompt because the speaker abdicates control.” (EX-1013, 56.) To put it simply, what Whittaker and Stenton established was that the best way for a speaker to maintain control in dialogue is through directives (commands).

### **iii. Conversational**

73. A third configuration for spoken dialogue systems is a “conversational” approach. As Möller wrote “[a] main characteristic of the conversation model is the distribution of initiative between the system and the user.” (EX-1008, 30.) Another characteristic of a conversational approach is a focus on collaboration. As Möller explains: “Understanding can be reached either by a collaborative interaction between system and user, making use of questions and confirmation (the approach which is of interest for SDSs), or by extracting information without any feedback or interaction (e.g. text summarization).” (EX-1008, 115.) In a collaborative approach, “[i]nstead of concentrating on the structure of the task (as in plan-based approaches),” the system tries “to capture the motivation behind a dialogue, and the dialogue mechanisms themselves.” (EX-1008, 28.)

74. While truly “conversational” systems are ostensibly the most capable and adaptable type of dialogue systems, they are not without trade-offs. Möller summarizes this issue nicely:

It may appear obvious that users would prefer a more flexible interaction style, thus mixed-initiative dialogues. However, mixed-initiative [exploratory or conversational] dialogues are generally more complex, in that they require more knowledge on the part of the user about the system capabilities. The possibility to take the initiative leads to longer and more complex user queries which are more difficult to recognize and interpret. Consequently, more errors and correction

dialogues might impact the user's overall impression of a mixed-initiative system. This observation has been made in the evaluation of the ELVIS E-mail reader system by Walker et al. (1998a), where the mixed initiative system version – although being more efficient in terms of the number of user turns and the elapsed time to complete a task – was less preferred by the users against a system-initiative version. It was assumed that the additional flexibility caused confusion for the users about the possible options, and lead to lower recognition rates.

(EX-1008, 30.)

75. However, as of the earliest priority date of the '765 patent, a number of such systems had been developed, primarily by academic research centers across the United States and in Europe. In the United States, these included the DARPA Communicator systems, the TRAINS and TRIPS systems developed by the Conversational International and Spoken Dialogue Research Group at the University of Rochester, and the Conversational Architectures Project at Microsoft. (EX-1009, 323.)

## **B. Overview of the '765 Patent**

76. The '765 patent claims priority to non-provisional U.S. Application No. 11/580,926, filed on October 16, 2006. Although I do not have an opinion about whether the '765 patent is entitled to this priority date, each applied reference was filed or published before October 16, 2006.

77. The '765 patent is titled “System and Method for a Cooperative Conversational Voice User Interface” and is generally directed to “a cooperative conversational model for a human to machine voice user interface.” (EX-1001, 1:17-18.)

78. Each of the claims of the '765 patent requires at minimum three elements: (1) a “user”, (2) a “voice-enabled device”, and (3) “a speech engine.”

79. Figure 1 of the '765 patent, which I reproduce below, shows an “exemplary system architecture” for implementing the conversational voice user interface. (EX-1001, Fig. 1., 7:14-17.)

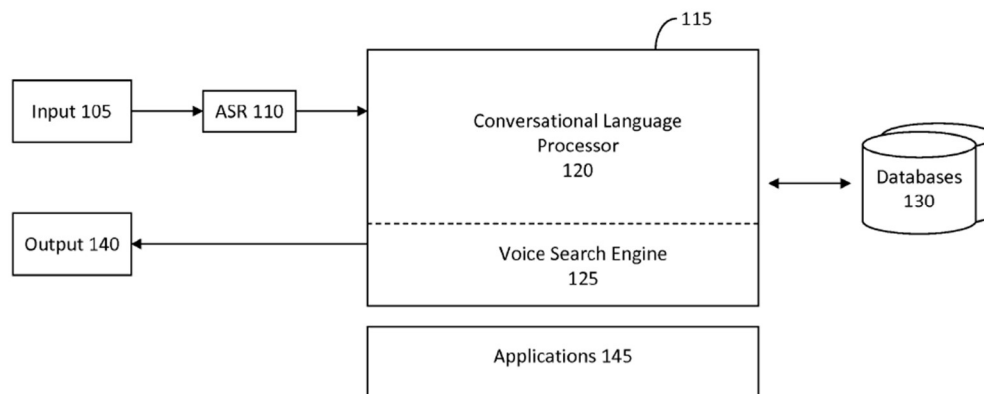


**Figure 1**

### **'765 Patent, Figure 1**

80. As shown in Figure 1 of the '765 patent, the “utterance component of input 105 may be processed by a speech recognition engine 110 ... to generate one

or more preliminary interpretations of the utterance. (EX-1001, 7:31-35.) “The one or more preliminary interpretations generated by the speech recognition engine 110 may then be provided to a conversational speech engine 115 for further processing.” (EX-1001, 7:44-47.)

81. Most relevant to the claims at issue here is Figure 3 of the '765 patent, reproduced below. That depicts an “exemplary block diagram of a cooperative conversational model, according to one aspect of the invention.” (EX-1001, 7:7-9.) The specification provides that “a cooperative conversation may take incoming data (shared knowledge) 305 to inform a decision (intelligent hypothesis building) 310, and then may refine the decision and generate a response (adaptive response building) 315.” (EX-1001, 13:34-38.)

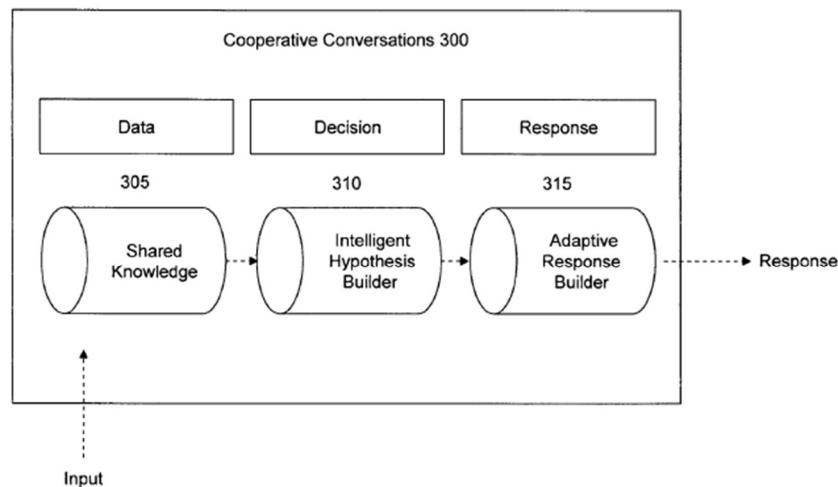


FIG. 3

**'765 Patent, Figure 3**

82. As shown in Figure 3, an adaptive response is then constructed in operation 315, which “may build syntactically, grammatically, and contextually sensitive ‘intelligent responses’ that ... guid[e] the user to reply in a manner favorable for recognition.” (EX-1001, 17:26-31.)

83. The adaptive response may vary tone, grammar, or formality, and “may frame responses to influence a user reply utterance for easy recognition,” including offering verbal options or rephrasing prior user language. (EX-1001, 18:18-43.)

84. Figure 2 of the '765 patent, which I reproduce below, shows components of the conversational speech engine 215, including a “free form voice search module 245,” “noise tolerance module 250,” and “context determination process 255.” (EX-1001, Fig. 2, 8:6-8.)

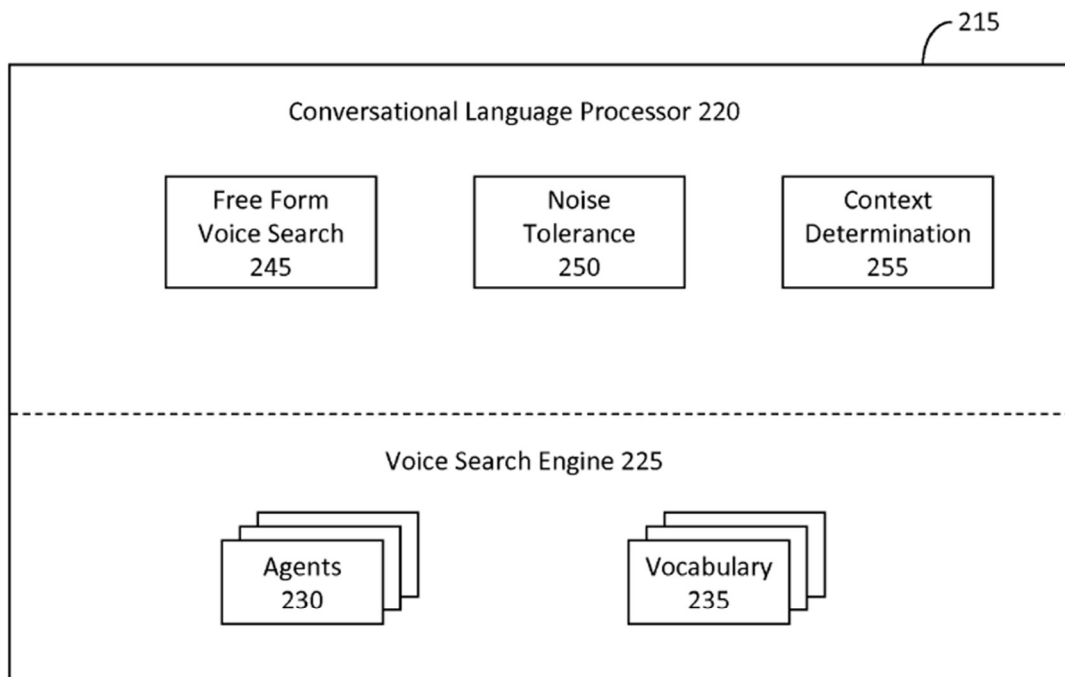


**Figure 2**

### **'765 Patent, Figure 2**

#### **C. Disclosures Relevant to the Claims of the '765 Patent**

85. All meaningful disclosures relevant to the claims of the '765 patent are found at 5:38-54 and 14:37-15:40.

86. The '765 patent describes that one component of a successful cooperative conversation involves identifying a type of conversation from a user utterance. (EX-1001, 14:38-40.) According to Figure 3, this classification occurs as part of the hypothesis-building operation 310, which includes identifying

“conversational goals, participant roles, and/or an allocation of information among the participants.” (EX-1001, 14:43-46.)

87. The conversational goals identified by the Intelligent Hypothesis Builder 310 may include “(1) getting a discrete piece of information or performing a discrete task, (2) gathering related pieces of information to make a decision, and/or (3) disseminating or gathering large amounts of information to build expertise.” (EX-1001, 14:46-50.)

88. Roles in the conversation may vary: the user may act as a “leader” (directing the dialogue), “supporter” (assisting), or “consumer” (receiving information), depending on the nature of the task and how the information is distributed. (EX-1001, 14:51-58.) Intelligent Hypothesis Builder 310 may adapt dynamically as goals shift, adjusting hypotheses accordingly. (EX-1001, 14:64-67.)

89. The ’765 patent categorizes conversations into three general types, with roles and information flow defined both in prose and in a tabular model in the specification, namely:

- **Query conversations:** The user (Participant A) acts as a leader/consumer, and the voice user interface (Participant B) acts as a supporter/dispenser. The user seeks specific information or a task to be performed. (EX-1001, 15:1-7, 15:25-28.)

- **Didactic conversations:** The voice interface takes the lead, guiding the user (*e.g.*, in instructional or tutorial contexts), while the user acts as a passive recipient. (EX-1001, 15:7-12, 30-32.)
- **Exploratory conversations:** Both user and interface may alternate roles, jointly shaping the conversation without a fixed goal. (EX-1001, 15:12-16, 35-39.)

90. Notably, outside these disclosures about the characteristics of query, didactic, and exploratory conversations, the patent does not disclose any specific algorithm or structural mechanism for identifying which type of conversation type an utterance belongs to. Instead, it broadly describes a black-box “Intelligent Hypothesis Builder 310” that “may broadly categorize a conversation” using the parameters above, such as conversational goals, participant roles, and information allocation. (EX-1001 at 15:16-18.). However, it stops short of explaining *how* these classifications would actually be computed or operationalized in a system. There is *no* description of the logic, rules, statistical models, or machine learning techniques that would be used to distinguish or identify one conversation type from another. Even more, there is no description of the logic, rules, statistical models, or machine learning techniques that would allow the system to assess the subsidiary parameters, such as how are the goals, roles, and information allocation variables determined based on a series of utterances.

#### D. Prosecution History

91. During prosecution, the Examiner rejected the claims as obvious over U.S. 2005/0015256 (“Kargman”) in view of U.S. 2007/0073544 (“Millett”).

(Prosecution History for U.S. 8,515,765; EX-1002, 106.)

92. Applicants filed a response on January 19, 2012, withdrawing claims 1-14, amending claims 15-26 to attempt to overcome the obviousness rejection, and adding new claims 27 and 28. For example, pending independent claim 15 was amended to add the limitations shown below, where the added limitations are underlined and the deleted limitations are struck through:

15. **(Currently Amended)** A method ~~for~~ of providing a ~~cooperative conversational~~ voice user interface, comprising:

~~receiving~~ receiving ~~[[an]]~~ a natural language utterance ~~[[at]]~~ from a voice input device during a ~~current~~ conversation between the voice user interface and a user;

determining ~~classifying~~, at a conversational speech engine, a conversation type of the ~~current~~ conversation or of the natural language utterance received during the ~~current~~ conversation, wherein determining the conversation type includes determining a conversational goal of the conversation into a conversation type;

~~establishing an intended meaning associated with the utterance based on the conversation type associated with the current conversation or the utterance; and~~

generating a grammatically or syntactically adapted response to the natural language utterance based on the conversation type ~~intended meaning associated with the utterance and the conversation type.~~

(EX-1002, 127-28.)

93. In accompanying remarks, Applicants argued that Millett failed to teach a “response [that] is grammatically or syntactically adapted” because Millett “relates to determining whether a variable in a resource string should be in a singular or in a plural form” but “is in no way related to (nor based on) a conversation type.” (EX-1002, 134-35.) Applicants claimed that Millett does not disclose “how it would be obvious for the grammatically or syntactically adapted response to be based on a conversational goal or conversation type.” (EX-1002, 135.)

94. On August 6, 2012, the Examiner issued a Final Rejection, maintaining the obviousness grounds. The Examiner explained that the Applicants appeared to be arguing for an interpretation of the claims that was “not the only interpretation of the claim language and is a feature not required as claimed.” (EX-1002, 152.) The Examiner further explained that “all that is required in the claims is a ‘response to the natural language utterance’ that has the characteristics of being ‘grammatically or syntactically adapted’ and which was ‘generated’ ‘based on conversation type.’ The generation of the response to the natural language utterance is not limited to only the grammatical or syntactic adaptation, and can include, among other things the selection of what words are in the response.” (EX-1002, 152-53 (emphasis in original).)

95. Applicants initiated a telephonic interview on August 9, 2012. (EX-1002, 182.) According to Applicants, the Examiner indicated in the interview that

“claims 16, 27, and 28 would be allowable if rewritten in independent form to include all of the features of the base claim and any intervening claims.” (EX-1002, 182.)

96. Applicants filed a Request for Continued Examination on November 27, 2012, cancelling claims 1-15 and 21, amending claims 16-20 and 22-28 to attempt to overcome the obviousness rejection, and adding claims 29 and 30. For example, pending claim 16 was amended to add the limitations shown below, where the added limitations are underlined and the deleted limitations are struck through:

16. **(Currently Amended)** ~~The method recited in claim 15,~~ A method of providing a voice interface, comprising:  
receiving, at a speech engine, a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type;  
determining the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance ~~wherein determining the conversation type further comprises determining whether the user~~ has a leader role in an interaction with the voice-enabled device in the conversation ~~or has a supporter role in the interaction with the voice-enabled device in the conversation;~~ and  
generating a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the ~~, and wherein generating the grammatically or syntactically adapted response comprises generating, if the user has a supporter role in the conversation, a response that limits a user's future input to interjecting queries or requests for clarification~~ if the user has the supporter role.

(EX-1002, 177.)

97. In accompanying remarks, Applicants highlighted the new feature of claim 16, which required “generating a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user’s future input to interjecting queries or requests for clarification if the user has a supporter role.” (EX-1002, 183.) Applicants argued that Kargman failed to disclose “generating a response with a format based on conversation type nor using a format that limits a user’s future input to interjecting queries or requests for clarification if the user has a supporter role.” (EX-1002, 184.)

98. The Examiner allowed claims 16-20 and 22-30 on February 14, 2013. (EX-1002, 197.) The Examiner stated that “the prior art of record does not teach the limitations previously indicated as allowable subject matter in the Final Office Action mailed 8/6/12.” (EX-1002, 197.) Applicants requested continued examination to allow the Examiner to consider an additional IDS, and the Examiner subsequently amended the Notice of Allowance to indicate that the IDS had been considered. (EX-1002, 220.)

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

99. I understand that certain issues relating to the validity of the ’765 patent must be judged from the perspective of a person of ordinary skill in the relevant art, as I discuss below. I have been asked to define the level of a “person of ordinary skill in the art” or “POSITA” at the time the alleged invention as claimed was made.

In deciding the level of ordinary skill, I have considered the following factors, which I have been informed are relevant to the determination of ordinary skill in the art.

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

100. In my opinion, a person of ordinary skill in the art would have a bachelor's degree in computer science, computer engineering, electrical engineering, or related field in computing technology, and two years of experience with automatic speech recognition and natural language understanding, or equivalent education, research experience, or knowledge.

#### **F. Claim Construction**

101. I understand that “claim construction” is the process of determining a patent claim's meaning. I understand that the Patent Office applies the same claim construction standard used in district courts which seeks to give claim terms their plain and ordinary meaning to a POSITA at the time of the claimed invention in view of the claims, specification, and prosecution history. I also understand that if the patent applicant gave a term a special meaning in the specification or during prosecution, then the special meaning should be used.

#### **IV. GROUND 1: COMBINATION OF SMITH & HIPPI AND GUINN RENDERS OBVIOUS CLAIMS 1-14.**

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

102. The combination of “An Architecture for Voice Dialog Systems Based on Prolog-Style Theorem Proving” to R.W. Smith, et al. (“Smith & Hipp”; EX-1005) and “Mechanisms for Mixed-Initiative Human-Computer Collaborative Discourse” to C. Guinn (“Guinn”; EX-1006) discloses every limitation of claims 1-14 of the ’765 patent. Smith & Hipp and Guinn are analogous art to the ’765 patent because both references are directed to speech recognition systems and software. *See also* (EX-1007.)

103. I describe Smith & Hipp and Guinn further below.

##### **1. Overview of Smith & Hipp**

104. Smith & Hipp (1995) discloses a voice dialog system architecture built around a Prolog-style theorem prover that enables efficient, adaptive human-machine interaction. (EX-1005, 281.) The system is designed to support real-time dialog in task domains such as electronic circuit repair and is capable of generating responses, managing user input, switching between dialog threads, and modeling user knowledge.

105. The dialog system’s central mechanism is a Prolog-style inference engine, which Smith & Hipp called the Interruptible Prolog SIMulator (“IPSIM”) theorem prover. (EX-1005, 287.) The IPSIM theorem prover attempts to satisfy a

task goal using internal knowledge. If the goal cannot be resolved internally, the system engages in dialog to obtain “missing axioms” from the user. (EX-1005, 287.) The authors explain that “our system is built around a theorem prover at its core, and the role of language is to supply missing axioms.” (EX-1005, 286.)

106. The system architecture supports behavior “required for efficient human-machine dialog,” including:

- (1) problem solving to achieve a target goal;
- (2) the ability to carry out subdialogs to achieve appropriate subgoals and to pass control arbitrarily from one subdialog to another;
- (3) the use of a user model to enable useful verbal exchanges and to inhibit unnecessary ones;
- (4) the ability to change initiative from strongly computer controlled to strongly user controlled or to some level in between;
- (5) the ability to use context dependent expectations to correct speech recognition and track user movement to new subdialogs.”

(EX-1005, 281.)

107. Smith & Hipp discloses that the dialog system supports variable initiative, meaning (as explained before) the degree of control over the dialog can shift between the computer and the user. The system is capable of operating in different initiative modes ranging from strongly system-directed to strongly user-directed, with intermediate levels in between. (EX-1005, 282-83.)

108. Specifically, the system defines four levels of dialog initiative or dialog control, which I describe in the following paragraphs.

109. In a fully **directive** mode, the computer leads the interaction at each step: “The computer has complete dialog control. It recommends a subgoal for completion and will use whatever dialog is necessary to obtain the needed item of knowledge related to the subgoal.” (EX-1005, 282.) In this mode, the system proactively guides the user through the troubleshooting procedure, deciding which question or instruction to give next and only relinquishing control if absolutely needed.

110. At the opposite extreme, a **passive** mode is defined in which “[t]he user has complete dialog control. The computer responds directly to user questions and passively acknowledges user statements without recommending a subgoal as the next course of action.” (EX-1005, 283.) In **passive** mode, the system will not volunteer any new directives or hints; it simply waits for the user to ask questions or state observations, and it responds or confirms accordingly.

111. Between these extremes lie intermediate modes such as **suggestive** and **declarative**. In **suggestive** mode, the system still predominantly controls the dialog but in a less forceful manner – it may propose the next subgoal or action rather than command it, and it allows the user to make minor digressions or clarifications related to the task at hand. (EX-1005, 283, 289.) In declarative mode, the user generally

drives the dialog (able to introduce new subdialogs at will), but the system remains cooperative by adding relevant information or reminders unprompted, if they might help the user. Essentially, declarative mode lets the user take initiative while the system acts as a knowledgeable assistant, chiming in with useful facts when appropriate. (EX-1005, 283, 289.)

112. The Smith & Hipp architecture adjusts several aspects of the dialog behavior based on the current mode. One mechanism is through the dialog controller's control over subdialog interruptions: in a more system-driven mode, the system will restrict the user's ability to deviate to new topics (only allowing interruptions for clarifications), whereas in user-driven modes, the controller will permit the user to initiate entirely new subdialogs freely. (EX-1005, 289.)

113. These initiative levels (sometimes called initiative modes) affect how the system formulates responses in order to maintain initiative. As Smith & Hipp describes, "[t]he system must output statements that are compatible with its level of initiative. This means that the output generator must have a parameter that enables the system to specify assertiveness." (EX-1005, 289.) For example, the same command may be phrased "at various levels of assertiveness" depending on the mode:

|                                        |                     |
|----------------------------------------|---------------------|
| "Turn the switch up."                  | (command)           |
| Would you [please] turn the switch up? | (request)           |
| Can you turn the switch up?            | (indirect request)  |
| The switch can be turned up.           | (statement of fact) |

Turning the switch up is necessary. (statement of fact)

(EX-1005, 289.)

114. Put slightly differently, to communicate the same intended action, a **directive** mode prompt might be a firm command (“Turn the switch up.”), a **suggestive** mode prompt might be phrased as a polite request (“Would you please turn the switch up?”), and a more user-controlled mode might even state it as an observation or necessity (“The switch can be turned up,” or “Turning the switch up is necessary.”) (EX-1005, 289.)

115. By varying the linguistic assertiveness (command vs. question vs. statement of fact), the system signals how strongly it is taking the lead in the task. Smith & Hipp notes that even silence handling is mode-dependent: the system tolerates longer pauses from the user in **passive** mode than it would in **directive** mode, since in **passive** mode it expects the user to take initiative (EX-1005, 290.) All these adjustments are geared toward making the dialog feel natural for the chosen interaction style.

116. The mode of initiative influences the system’s behavior in selecting the next subgoal, shaping system prompts, and interpreting user input. In **directive** mode, for instance, the system assumes the user has little task information and leads the dialog step-by-step; in declarative mode, the user leads the interaction, and the system contributes only supportive information. (EX-1005, 288-90.)

117. In addition to initiative levels, the system maintains an explicit user model, which is continuously updated throughout the interaction. The system continually updates a set of Prolog-style facts (rules) reflecting what the user knows, can do, or has done so far in the dialog. This user model is dynamically updated because, as Smith & Hipp notes, “[a]lmost every statement will change the user’s knowledge base, and future interactions will be ill-conceived if the appropriate updates are not made.” (EX-1005, 288.)

118. This model includes representations of what the user knows or is capable of, and it affects whether the system explains a step, asks a clarifying question, or simply proceeds. (EX-1005, 287-88.) For example, if a user asks for clarification on a task (*e.g.*, “Which knob?”), the system will detect a mismatch with its existing user model, revise the model, and insert a clarification subdialog accordingly. (EX-1005, 296-97.)

119. This user model also is used to affect the initiative mode of the system. As Smith & Hipp explains: “A real possibility in a cooperative interaction is that the user’s problem-solving ability, either on a given subgoal or on the global task, may exceed that of the machine. When this occurs, an efficient interaction requires that the machine yield control so that the more competent partner can lead the way to the fastest possible solution. Thus, the machine must be able to carry out its own

problem-solving process and direct the actions to task completion or yield to the user's control and respond cooperatively to his or her requests.” (EX-1005, 282.)

120. The system also uses dialog context to generate expectations about what a user is likely to say next. These expectations are categorized as situation-specific, task-related, and dialog-structure-related, and they are used both to guide recognition and to detect when the user shifts from one subgoal or conversation segment to another. (EX-1005, 290-92.) For example, after issuing a task command, the system might expect short confirmations, requests for clarification, or comments on related subtasks. These expectations help guide the speech recognizer and reduce ambiguity in input interpretation. (EX-1005, 291-92.)

121. Smith & Hipp further describes the system's ability to interpret utterances based on the user's role in the conversation. The system distinguishes between situations where the user is providing information, requesting information, or following instructions. (EX-1005, 290-91.) These dialog patterns, combined with conversational goals, help the system shape its responses appropriately—for instance, by delivering a short answer, a fuller explanation, or requesting information in return. (EX-1005, 291.)

122. Smith & Hipp conducted a series of experiments to evaluate the dialog system's performance in **directive** and **declarative** initiative modes using eight human subjects tasked with diagnosing faults in an electronic circuit through natural

language dialog. (EX-1005, 303-304.) The results demonstrated that “the completion times and numbers of utterances were far smaller for the **declarative** mode.” (EX-1005, 308.) **Declarative** mode produced more concise and natural exchanges, while **directive** mode resulted in “far more questions from the machine with only one-word answers,” leading to “faster responses and higher recognition rates” but greater dialog length overall. (EX-1005, 308.) Notably, the study confirmed that dialog behavior adapted to the user’s role and level of control: “During session 1, subjects had gained substantial information about how to debug this circuit, so the extremely pedantic **directive** mode was too detailed for them in most cases,” whereas “they functioned much better in **declarative** mode, in which the machine stated relevant facts but allowed the user to lead the interaction.” (EX-1005, 308.)

## 2. Overview of Guinn

123. Guinn was published in 1996 as a follow-on to the voice dialog system developed by Smith & Hipp. Guinn’s Abstract states:

In this paper, we examine mechanisms for automatic dialogue initiative setting. We show how to incorporate initiative changing in a task-oriented human-computer dialogue system, and we evaluate the effects of initiative both analytically and via computer-computer dialogue simulation.

(EX-1006, 278.)

124. Guinn recognized that the Smith & Hipp system did not include automatic initiative switching and sought to provide a model for automatic switching: “The dialogue model of Smith (1995) allows for either the computer or the user to assume degrees of initiative; however, Smith presents no algorithm for the computer to change initiative during a dialogue. Our model of mixed-initiative dialogue allows either participant to be in control of the dialogue at any point in time.” (EX-1006, 278.)

125. Guinn discloses that, “[i]n our model of dialogue, initiative levels for each goal are defined during the interaction based on 1) explicit and implicit initiative-changing utterances and 2) competency evaluation.” (EX-1006, 279.)

126. With respect to explicit and implicit initiative-changing utterances, Guinn’s model “concentrates on two specific dialogue cues: questions and answers.” (EX-1006, 279.) For example, “When an agent  $A_1$  asks another agent  $A_2$  to satisfy a goal  $G$ , agent  $A_2$  gains initiative over goal  $G$  and all subgoals of  $G$  until agent  $A_2$  passes control of one of those subgoals back to agent  $A_1$ .” (EX-1006, 279 (emphasis in original).)

127. As for competency evaluation, Guinn’s methodology “involves a probabilistic examination of the search space of the problem domain.” (EX-1006, 279.) According to Guinn, “[c]ompetency evaluation will be based on how likely an agent’s branch will be successful (based on a weighted factor analysis) and how

likely the collaborator's branch will be successful (based on a weighted factor analysis and a probabilistic model of the collaborator's knowledge)." (EX-1006, 279.)

128. Guinn investigated four initiative-setting schemes, using these two key factors:

- **“Random** In Random mode, one agent is given initiative at random in the event of a conflict. This scheme provides a baseline for initiative setting algorithms. Hopefully, a proposed algorithm will do better than Random.” (EX-1006, 281.)
- **“SingleSelection** In SingleSelection mode, the more knowledgeable agent (defined by which agent has the greater total percentage of knowledge) is given initiative. The initiative is set throughout the dialogue. Once a leader is chosen, the participants act in a master-slave fashion.” (EX-1006, 281.)
- **“Continuous** In Continuous mode, the more knowledgeable agent (defined by which agent's first-ranked branch is more likely to succeed) is initially given initiative. If that branch fails, this agent's second-ranked branch is compared to the other agent's first-ranked branch with the winner gaining initiative. In general if Agent 1 is working on its *i*th-ranked branch and Agent 2 is working on its *j*th-ranked branch, we compare  $p^{A_1}(b_i^{A_1})$  to  $p^{A_2}(b_j^{A_2})$ .” (EX-1006, 281.)
- **“Oracle** In Oracle mode, an all-knowing mediator selects the agent that has the correct branch ranked highest in its list of branches. This scheme is an upper bound on the effectiveness of initiative setting schemes. No initiative setting algorithm can do better.” (EX-1006, 281.)

129. Guinn concludes that “SingleSelection and Continuous modes perform significantly better than Random mode. On average Continuous mode results in 40% less branches searched per goal than Random. Continuous mode performs between 15-20% better than SingleSelection.” (EX-1006, 281-82.)

130. My analysis in this Declaration focuses on the Continuous mode disclosed in Guinn.

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

131. The claims of the '765 patent require determining a conversation type (*i.e.*, an initiative mode in Smith & Hipp) *based on* whether the “role” of the user in the conversation. As explained above, Smith & Hipp, like the '765 patent, teaches different conversation types that are based on user roles. For example, when the user is in a leader role, Smith & Hipp teaches utilizing the “passive” conversation type. Alternatively, when the user is in a completely supportive role, Smith & Hipp teaches the “directive” conversation type.

132. Smith & Hipp states that a primary objective of its system is allow continuous updating of roles and, in turn, conversation types. For example, Smith & Hipp states: “Because considerable information is exchanged during the dialog, the user model changes continuously.” (EX-1005, 282).

133. While Smith & Hipp advocates for this operational dynamic in theory, it also states that the system as implemented by the authors needed “an automatic

mechanism ... for setting the appropriate level of initiative as an interaction proceeds.” (EX-1005, 316). Thus, Smith & Hipp discloses automatically switching modes in theory in the same way as the ’765 patent, which for itself, includes no actual algorithms teaching a POSITA how to automatically change conversation types based on shifting roles.

134. To the extent an actual implementation, instead of Smith & Hipp’s well described theory, showing automatic initiative switching is required by the claims, a POSITA would have been motivated to utilize the teachings of Guinn (particularly using Guinn’s Continuous mode) in the Smith & Hipp system, particularly to resolve Smith & Hipp’s urgings that advancement was needed to even further accomplish the goal of switching between initiative modes as the relative role of the user and system were updated based on necessary control.

135. A POSITA would have been motivated to make this combination for numerous reasons. First, as explained above, Smith & Hipp expressly suggests to readers and motivates them by identifying an “important unsolved problem[,]” which is achieving the described automatically switching between initiative modes: “A number of important unsolved problems need attention to enable further progress. For example, **an automatic mechanism is needed for setting the appropriate level of initiative as an interaction proceeds.**” (EX-1005, 316.) A

POSITA would therefore have been motivated look to other references to address this important unsolved problem in Smith & Hipp's system.

136. Second, Guinn latches onto this motivation and emphasizes that a POSITA would have been motivated to consider and implement its teachings in the Smith & Hipp system: "The dialogue model of Smith (1995)<sup>1</sup> allows for either the computer or the user to assume degrees of initiative; however, **Smith presents no algorithm for the computer to change initiative during a dialogue.**" (EX-1006, 278). Notably, Guinn goes on to confirm that its teachings resolve the implementation problem identified by Smith & Hipp: "Our model of mixed-initiative dialogue allows either participant to be in control of the dialogue at any point in time." (EX-1006, 278.)

137. Third, the combination is nothing more than the application of a known technique (automatic initiative switching in a voice dialog system as disclosed in Guinn) to improve a known system (Smith & Hipp's voice dialog system) that operates in the same way (setting of initiative modes in conversations based on the relative knowledge of user and computer).

138. Fourth, the combination would have been achieved through known methods (utilizing the algorithm structure disclosed by Guinn in a voice dialog

---

<sup>1</sup> I note that Guinn was referencing a different reference published by Smith and Hipp than the article used in Ground 1 of this petition. However, both articles concerned the same underlying spoken dialogue system.

system) to achieve predictable results (automatic initiative switching, which is promoted in theory by Smith & Hipp).

139. Fifth, Smith & Hipp expressly states that its system is ready for improvement, considering its disclosure that while automatic mode switching is desired in theory, it has not yet been implemented. This is a plain statement that the system is ready for improvement. Guinn is particularly well-suited as a known technique for providing predictable results considering its expressly teaches how to solve the problem.

140. A POSITA would have been motivated to look to the teachings of Guinn.

141. A POSITA also would have found the combination – updating the Smith & Hipp system to achieve automatic initiative switching – easily achievable considering Guinn expressly discloses that part of its goal is focused on improving this aspect of Smith & Hipp.

142. As such, a POSITA would have had a reasonable expectation of success in the combination and the results of the combination would have been predictable. Combining Smith & Hipp with Guinn would merely integrate one component (automatic switching of initiative modes as disclosed in Guinn) into Smith & Hipp's system of receiving and responding to user's utterances. As already noted above, Smith & Hipp and Guinn both expressly recognized that Smith & Hipp did not

include, but could benefit from, automatic switching of initiative modes during a conversation. (EX-1005, 316; EX-1006, 278.) Implementing Guinn’s automatic switching—which was expressly intended to build upon the work of Smith & Hipp—into the Smith & Hipp system would have been well within the skill set of a POSITA before the earliest priority date of the ’765 patent.

143. Smith & Hipp and Guinn are both in the same field of endeavor as the ’765 patent—providing a cooperative conversational voice user interface. (*See, e.g.*, EX-1001, Title (“System and Method for a Cooperative Conversational Voice User Interface”), 1:17-18 (“The invention relates to a cooperative conversational model for a human to machine voice user interface.”); EX-1005, 281 (“This paper presents a theory of voice dialog systems that integrates into a single, self-consistent architecture a variety of capabilities necessary for successful dialog. These capabilities include problem solving, natural language input and output, a user model, variable initiative, and the use of expectation for speech error correction and plan recognition.”), 282 (“The purpose of the architecture is to deliver to users in real time the behaviors needed for efficient human-machine dialog.”); EX-1006, 278 (“We show how to incorporate initiative changing in a task-oriented human-computer dialogue system ... ”).)

## **B. Independent Claims 1 and 10**

144. Independent claims 1 and 10 include substantially overlapping

processing limitations ([1A]-[1C]/[10A]-[10C]), as shown in the table below. The only substantive difference between the claims is that claim 1 claims a method for selecting and presenting advertisements and claim 10 claims a system that practices that same method for selecting and presenting advertisements. Because of the overlap of the limitations in these claims, I address these limitations together. In this table, I also provide a citation to the section discussing each limitation:

| <b>Claim 1</b>                                                                                                                                                                                                                                                                 | <b>Claim 10</b>                                                                                                                                                                                                                                                                                     | <b>Section</b> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| [1P] A method of providing a voice interface, comprising:                                                                                                                                                                                                                      | [10P] A system for providing a voice interface, comprising:                                                                                                                                                                                                                                         | §IV.B.1        |
| [1A] receiving, at a speech engine, a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type;                                                                                                             | [10A] a speech engine configured to receive a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type, wherein the speech engine includes a processor configured to:                                                            | §IV.B.2        |
| [1B] determining the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the | [10B] determine the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device; | §IV.B.3        |

|                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                     |         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| voice-enabled device;<br>and                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                     |         |
| [1C] generating a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user's future input to interjecting queries or requests for clarification if the user has the supporter role. | [10C] a response builder configured to generate a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user's future input to interjecting queries or requests for clarification if the user has the supporter role. | §IV.B.3 |

### 1. Preamble [1P] / Preamble [10P]

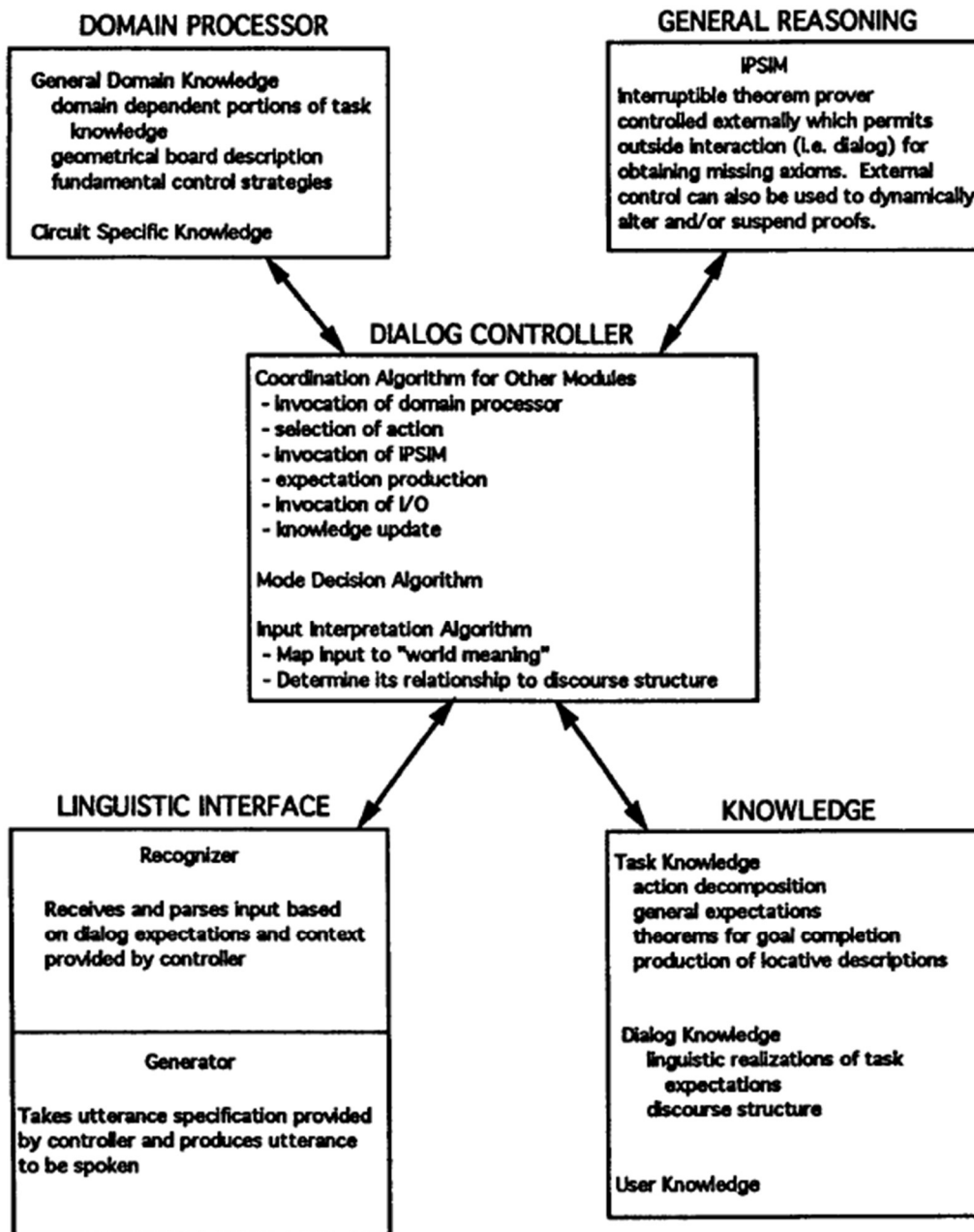
[1P] A method of **providing a voice interface, comprising:**

[10P] A system for **providing a voice interface, comprising:**

145. To the extent the preamble is limiting, the combination of Smith & Hipp and Guinn renders obvious “[a] *method of providing a voice interface*” as well as “[a] *system for providing a voice interface*,” as discuss below in §§IV.B.2-4. Smith & Hipp discloses a system that is used (*i.e.*, it performs a method) to provide a voice interface. Smith & Hipp states, “[a] pragmatic architecture for voice dialog machines aimed at the equipment repair problem has been implemented.” (EX-1005, 281.)

146. Smith & Hipp explains, “[t]he architecture of the [voice dialog] system is given in Figure 2, where five major subsystems are shown: the dialog controller,

the domain processor, the knowledge base, the general reasoning system, and the linguistic interface.” (EX-1005, 297.)



Smith & Hipp, Figure 2 (298)

147. At a high level, the five subsystems of the voice dialog system serve the following functions. I will address relevant functions in more detail in my analysis below.

- “*Dialog controller.* This is the overall ‘boss’ of the dialog processing system. It formulates goals at the top level to be passed on to the theorem-proving stage. It determines the role that the computer plays in the dialog by determining how user inputs relate to already established dialog as well as determining the type of response given. It also maintains all dialog information shared by the other modules and controls their activation. Its control algorithm is the highest-level dialog processing algorithm.” (EX-1005, 297.)
- “*The domain processor.* This is the primary application-dependent portion of the system. It contains much of the information about the application domain. It receives from the controller a request for the next suggested goal to be undertaken, and it returns to the controller its suggestion along with expected results from attempting the test. It then receives the results of the interaction and appropriately updates its data structures.” (EX-1005, 298.)
- “*The reasoning system.* The IPSIM system receives as input goals to be proven and commands to start, stop, and furnish information. It yields as output theorems that are proven and status reports on the proof in progress. Its processing follows the usual mechanisms of Prolog-style theorem-proving, and it is modeled by the ZmodSubdialog routine given above.” (EX-1005, 300.)
- “*The knowledge base.* This is the repository of information about task-oriented dialogs. This includes the following general knowledge about the

- performance of actions and goals: (1) Knowledge about the decompositions of actions into substeps. (2) Knowledge about theorems for proving completion of goals. (3) Knowledge about the expectations for responses when performing an action.” (EX-1005, 300.)
- “*Linguistic interface*. This system receives the spoken inputs from the user and returns spoken outputs. ... The inputs to the voice input system for each utterance are the set of expectations from the dialog controller and the speech utterance from the user. The output from this system is a GADL [Goal and Action Description Language] meaning representation.” (EX-1005, 300.)

148. Smith & Hipp discloses the following process performed by the dialog controller of the voice dialog system:

- (1) Obtain suggested goal from the domain processor.
- (2) Based on the suggested goal and the current state of the dialog, select the next goal to be pursued by the computer and determine the expectations associated with that goal. (The goal may thus be selected from one of the active subdialogs. The choice is partially dependent on the current level of initiative.)
- (3) Attempt to complete the goal using the IPSIM system, possibly invoking voice interactions.
- (4) Update system knowledge based on efforts at goal completion.
- (5) Determine next operations to be performed by the domain processor in providing a suggested goal.
- (6) Go to step 1.

(EX-1005, 297.)

149. Accordingly, and as further discussed below, the combination of Smith & Hipp and Guinn discloses “[a] *method of providing a voice interface*” and “[a] *system for providing a voice interface.*”

## 2. Receiving Natural Language Utterance [1A] / [10A]

[1A] receiving, at a speech engine, a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type;

[10A] a speech engine configured to receive a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type, wherein the speech engine includes a processor configured to:

150. The combination of Smith & Hipp and Guinn discloses “*receiving, at a speech engine, a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type*” and “*a speech engine configured to receive a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

151. As used in the ’765 patent, a POSITA would understand a “*speech engine*” to be a computer program implemented by one or more software modules.

152. For example, Figure 1 of the ’765 patent, reproduced below, depicts a conversational speech engine 115, which “may include a conversational language

processor 120 and/or a voice search engine 125 .... Conversational speech engine 115 may communicate with one or more databases 130 to generate an adaptive conversational response, which may be returned to the user as an output 140.” (EX-1001, 7:47-53.) The ’765 patent further states that, “in various implementations, the speech recognition engine 110, conversational speech engine 115, and/or databases 130 may reside locally (e.g., on a user device), remotely (e.g., on a server), or a hybrid model of local and remote processing may be used (e.g., lightweight applications may be processed locally while computationally intensive applications may be processed remotely).” (EX-1001, 7:61-67.)

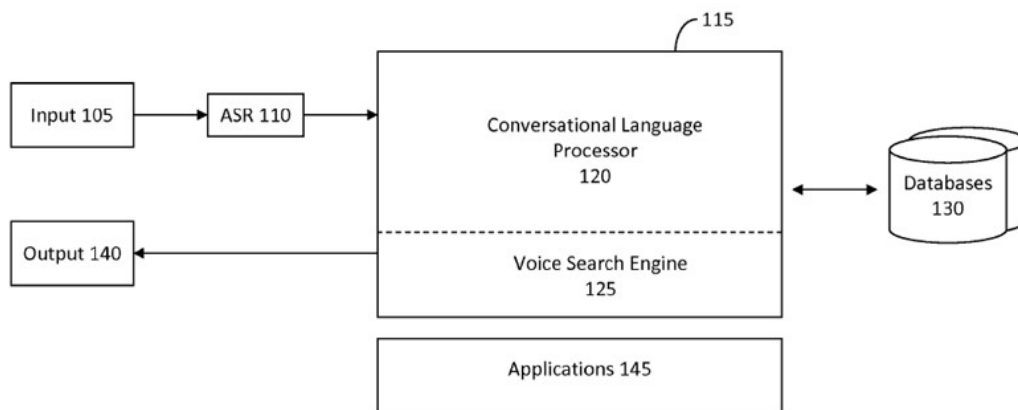
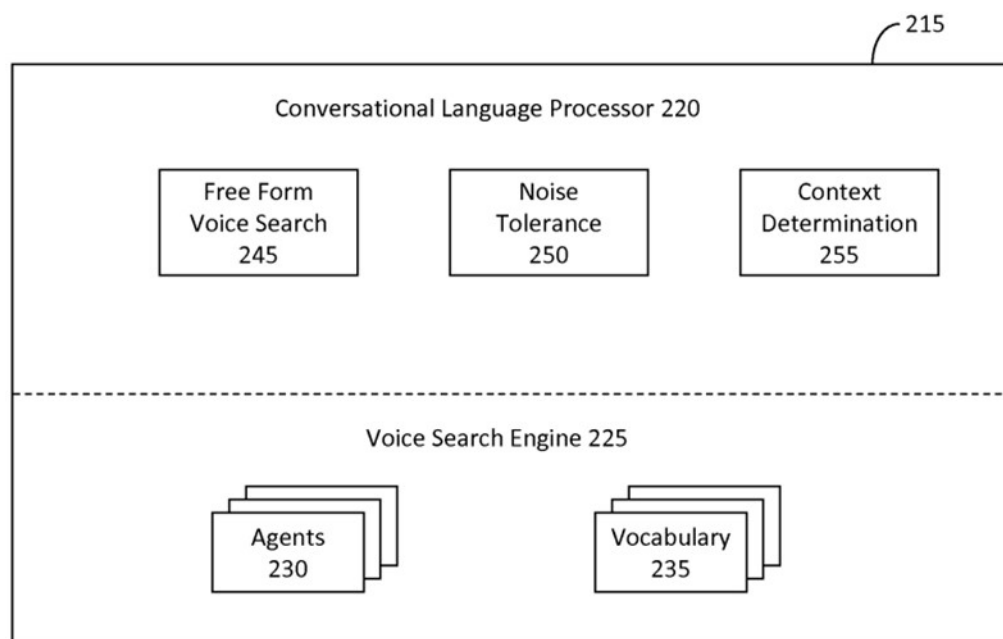


**Figure 1**

### **’765 Patent, Figure 1**

153. Figure 2 of the ’765 patent provides “an exemplary block diagram [] illustrating a conversational speech engine 215 according to one aspect of the

invention.” (EX-1001, 8:1-3.) “Conversational speech engine 215 may include a conversational language processor 220 that generates an adaptive conversational response to a request or series of requests using a free form voice search module 245, a noise tolerance module 250, and/or a context determination process 255.” (EX-1001, 8:3-8.)



**'765 patent, Figure 2**

154. According to the '765 patent, “an ‘utterance’ may be words, syllables, phonemes, or any other audible sound made by a human being.” (EX-1001, 2:16-18.)

155. A POSITA would have understood that, in the Smith & Hipp system, the combination of the five subsystems that I discussed above (dialog controller,

domain processor, reasoning system, knowledge base, and linguistic interface) together comprises a “speech engine” as that term is used in the ’765 patent.

156. According to Smith & Hipp, the dialog controller is one of five major subsystems implemented as software modules. (EX-1005, 297.) Smith & Hipp discloses:

*Dialog controller.* This is the overall “boss” of the dialog processing system. It formulates goals at the top level to be passed on to the theorem-proving stage. It determines the role that the computer plays in the dialog by determining how user inputs relate to already established dialog as well as determining the type of response given. It also maintains all dialog information shared by the other modules and controls their activation. Its control algorithm is the highest-level dialog processing algorithm.

(EX-1005, 297.)

157. Additionally, Smith & Hipp’s system includes a linguistic interface that receives voice inputs from users, also referred to as a “speech utterance” in Smith & Hipp:

*Linguistic interface.* This system receives the spoken inputs from the user and returns spoken outputs. ... The inputs to the voice input system for each utterance are the set of expectations from the dialog controller and the speech utterance from the user. The output from this system is a GADL [Goal and Action Description Language] meaning representation.

(EX-1005, 300.)

158. In order for the Smith & Hipp system to receive spoken inputs (speech utterances) from the user, a voice-enabled device, such as a microphone, must be used to receive the utterance from the user. (*See, e.g., EX-1001, 7:17-20* (“The system may receive an input 105 from a user, where in one implementation, input 105 may be an utterance received by an input device (e.g., a microphone), where the utterance may include one or more requests.”).) Indeed, Smith & Hipp discloses that “[s]peech recognition is performed by a Verbex 6000 user-dependent connected-speech recognizer running on an IBM PC.” (EX-1005, 303.) By necessity, the Verbex 6000 board application as described in Smith & Hipp would have included a voice-enabled device in the form of a microphone.

159. As shown in Figure 2 of Smith & Hipp (reproduced above), the linguistic interface receives the speech input from the user and, after some processing, passes the input to the dialog controller. Thus, the combination of Smith & Hipp and Guinn discloses “*receiving, at a speech engine, a natural language utterance from a voice-enabled device*” and “*a speech engine configured to receive a natural language utterance from a voice-enabled device.*”

160. Additionally, Smith & Hipp discloses that its system capabilities include “natural language input and output,” (EX-1005, 281), which a POSITA would understand corresponds to “*a natural language utterance*” as that term is used

in the '765 patent. Further, a POSITA would understand that any utterance provided by a human user to the Smith & Hipp system could be categorized as one more types of conversation. Thus, Smith & Hipp also discloses the received natural language utterance *“corresponding to a conversation type.”*

161. Accordingly, the combination of Smith & Hipp and Guinn discloses *“receiving, at a speech engine, a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type”* and *“a speech engine configured to receive a natural language utterance from a voice-enabled device, the natural language utterance corresponding to a conversation type.”*

### **3. Determining the Conversation Type Corresponding to the Natural Language Utterance [1B] / [10B]**

[1B] determining the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device; and

[10B] determine the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device;

162. The combination of Smith & Hipp and Guinn discloses *“determining the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an*

*interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device” and “determine the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device.”* These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

163. As discussed above, the ’765 patent identifies the following three conversation types, which differ based on the respective conversational goals, user roles, and allocation of information between the user and the system:

|                        | Participant A<br>User                     | Participant B<br>Voice User Interface     |
|------------------------|-------------------------------------------|-------------------------------------------|
| QUERY                  |                                           |                                           |
| GOAL                   | Get information/action                    | Provide information/action                |
| ROLE                   | Leader/Consumer                           | Supporter/Dispenser                       |
| INFORMATION ALLOCATION | Less                                      | More                                      |
| DIDACTIC               |                                           |                                           |
| GOAL                   | Get information                           | Provide information                       |
| ROLE                   | Follower/Consumer                         | Leader/Dispenser                          |
| INFORMATION ALLOCATION | Less                                      | More                                      |
| EXPLORATORY            |                                           |                                           |
| GOAL                   | Gather/share information                  | Gather/share information                  |
| ROLE                   | Follower/Consumer and<br>Leader/Dispenser | Follower/Consumer and<br>Leader/Dispenser |
| INFORMATION ALLOCATION | Equal or alternating                      | Equal or alternating                      |

(EX-1001, 15:20-40.)

164. As also discussed above, Smith & Hipp discloses that its system can operate in four initiative modes; these modes, like the '765 patent's conversation types, are determined on a shifting spectrum based on whether the user has control (*i.e.*, the user is the leader, the system is the supporter) or the system has control (*i.e.*, the system is the leader, the user is the supporter). The conversation types from the '765 patent correspond almost exactly to the four initiative modes disclosed in Smith & Hipp, except that Smith & Hipp puts a finer point on the spectrum covered by what the '765 patent calls an "exploratory" conversation where the relative role of the user and the system are constantly shifting:

| <b>Control</b>         | <b>Smith &amp; Hipp Initiative Mode</b> | <b>'765 Patent Conversation Type</b>      |
|------------------------|-----------------------------------------|-------------------------------------------|
| Computer               | Directive                               | Didactic                                  |
| Both Computer and User | Suggestive                              | Exploratory, With Bias Towards the System |
| Both Computer and User | Declarative                             | Exploratory, With Bias Towards the User   |
| User                   | Passive                                 | Query                                     |

165. Directive Mode = Didactic Conversation: The '765 patent discloses that, "In a didactic conversation, a leader of the conversation may control information desired by a supporter of the conversation. The supporter's role may be limited to regulating an overall progression of the conversation and interjecting

queries for clarification.” (EX-1001, 15:7-10; *see also* EX-1001, claim 4 (“the conversation type ... is determined as a didactic conversation if the conversational goal is to gather related information”).) In a didactic conversation, the conversation is led by the system, and the user is in a supporter role. (*See, e.g.*, EX-1001, claim 5 (“the conversation type ... is determined as a didactic conversation if the user that spoke the natural language utterance has the supporter role”).)

166. Smith & Hipp discloses the same conversation characteristics with respect to its **directive** mode. In **directive** mode, “[u]nless the user explicitly needs some type of clarification, the computer will select its response solely according to its next goal for the task. If the user expresses need for clarification about the previous goal, this must be addressed first. No interruptions to other subdialogs are allowed.” (EX-1005, 289; *see also* EX-1005, 282 (“The computer has complete dialog control. It recommends a subgoal for completion and will use whatever dialog is necessary to obtain the needed item of knowledge related to the subgoal.”).) In other words, the system controls the conversation as the leader, and the user of the system takes on a supporter role.

167. Suggestive Mode = Exploratory Conversation With Bias Towards the System; Declarative Mode = Exploratory Conversation With Bias Towards the User:  
The ’765 patent discloses that, “[i]n an exploratory conversation, both participants share leader and supporter roles, and the conversation may have no specific goal, or

the goal may be improvised as the conversation progresses.” (EX-1001, 15:12-16; *see also* ’765 patent, claim 5 (“the conversation type ... is determined as an exploratory conversation if the user that spoke the natural language utterance has both the leader role and the supporter role”).)

168. Smith & Hipp’s **suggestive** and **declarative** modes are equivalent to the ’765 patent’s exploratory conversation type, as the user and the computer may each have a leader or supporter role in those modes.

169. In **suggestive** mode, “[t]he computer still has dialog control, but not as strongly. The computer will make suggestions about the subgoal to perform next, but it is also willing to change the direction of the dialog according to stated user preferences.” (EX-1005, 283.) Thus, in **suggestive** mode, the user and the system may alternate serving the leader and supporter roles, though ultimately this mode weighs in favor of the computer having control (*i.e.*, being the leader).

170. In **declarative** mode, “[t]he user has dialog control, but the computer is free to mention relevant, though not required, facts as a response to the user’s statements.” (EX-1005, 283.) In this mode, the user and the system again may alternate serving the leader and supporter roles, though ultimately this mode weighs in favor of the user having control (*i.e.*, being the leader).

171. Passive Mode = Query Conversation: According to the ’765 patent:

[A] query conversation may include a conversational goal of getting a discrete piece of information or performing a particular task, where a leader of the query conversation may have a specific goal in mind and may lead the conversation toward achieving the goal. The other participant may hold the information and may support the leader by providing the information.

(EX-1001, 15:1-7; *see also* EX-1001, claim 4 (“the conversation type is determined as a query conversation if the conversational goal is to gather discrete information or perform a discrete task”).) A query conversation is user-led. (*See, e.g.*, EX-1001, claim 5 (“the conversation type is determined as a query conversation if the user that spoke the natural language utterance has the leader role”).)

172. In Smith & Hipp’s passive mode, the dialog characteristics are the same as a query conversation, *i.e.*, the user leads the conversation, *e.g.*, by asking direct questions, and the system supports the user by providing responses to the user’s questions. For example, Smith & Hipp discloses that, in **passive** mode, “[t]he user has complete dialog control. Consequently, the computer will passively acknowledge user statements. It will provide information only as a direct response to a user question.” (EX-1005, 289; *see also* EX-1005, 283 (“The user has complete dialog control. The computer responds directly to user questions and passively acknowledges user statements without recommending a subgoal as the next course of action.”).)

173. The '765 patent requires that the conversation type be “determined ... based on” a user’s role in the conversation at any given time. In other words, as roles shift, so too should the conversation type that is being applied by the system. When the user is in control as the leader, the system should apply a conversational type that is adaptive to the user’s role, which in Smith & Hipp would be adoption of the “passive” initiative mode. Likewise, when the user transitions to a supporter role and the system is in control, the system should adopt the “directive” mode.

174. Smith & Hipp teaches this transition type of the claims in theory. For example, Smith & Hipp explains that a voice dialog system “can participate in variable initiative dialog if it properly manages (1) the selection of the current subdialog, (2) the level of assertiveness in its outputs, and (3) the interpretation of its inputs.” (EX-1005, 289.) Smith & Hipp explains that “[o]ur theory also shows how variable initiative is built into the same simple architecture.” (EX-1005, 284). Smith & Hipp includes an entire section (4.4) directed towards this variable initiative goal, which “allows either participant to have control, and it allows the initiative to change between participants during the exchange.” (EX-1005, 289.)

175. In covering significant detail of variable initiative, Smith & Hipp teaches more about “determining” conversation type “based on” user role than the specification of the '765 patent, which covers the theoretical discussion in a few short sentences. (EX-1001, 14:51-58 (“Participant roles may broadly include: (1) a

leader that controls a conversation, (2) a supporter that follows the leader and provides input as requested, and/or (3) a consumer that uses information.

Information may be held by one or more of the participants at the outset of a conversation, where a participant may hold most (or all) of the information, little (or none) of the information, or the information may be allocated roughly equally amongst the participants.”.)

176. However, while Smith & Hipp discloses the use of various initiative modes that correspond to the conversation types of the ’765 patent and expresses an ideal theory for varying between conversation types, Smith & Hipp also states that its already built system did not include an automatic mechanism for selecting among the variable initiative levels or modes. (*See* EX-1005, 316 (noting that “an automatic mechanism is needed for setting the appropriate level of initiative as an interaction proceeds”); *see also* EX-1006, 278.)

177. To the extent it is necessary, Guinn provides the actual implementation that Smith & Hipp stated it had not yet achieved. Guinn teaches a “Continuous” mode of initiative selection and switching, where “the more knowledgeable agent (defined by which agent’s first-ranked branch is more likely to succeed) is initially given initiative.” (EX-1006, 281.)

178. When using Guinn’s Continuous mode in combination with the initiative modes in Smith & Hipp, if the combined system determines that the

computer is the more knowledgeable agent, initiative (or control) is given to the computer, such that the system operates in Smith & Hipp's **directive** mode, where "[t]he computer has complete dialog control." (EX-1005, 282.)

179. On the other hand, if the combined system determines that the user is the more knowledgeable agent, initiative is given to the user, such that the system operates in Smith & Hipp's **passive** mode, where "[t]he user has complete dialog control." (EX-1005, 283.)

180. In other words, implementing Guinn's teachings within the bounds of Smith & Hipp's disclosures, allows Smith & Hipp's implementation to achieve its goal of variable initiatives. In the combined system of Smith & Hipp and Guinn, the conversation type (or initiative mode) is determined based on whether the user ("*a user that spoke the natural language utterance*") has more knowledge (and thus the user is the leader of the conversation) or whether the computer has more knowledge (and thus the user is the supporter of the conversation).

181. Accordingly, Smith & Hipp discloses "*determining the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device*" and "*determine the conversation type corresponding to the natural language utterance based on whether a user that spoke the natural language*

*utterance has a leader role in an interaction with the voice-enabled device or has a supporter role in the interaction with the voice-enabled device.”*

**4. Generating a Response to the Natural Language Utterance  
[1C] / [10C]**

[1C] generating a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user’s future input to interjecting queries or requests for clarification if the user has the supporter role.

[10C] a response builder configured to generate a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user's future input to interjecting queries or requests for clarification if the user has the supporter role.

182. The combination of Smith & Hipp and Guinn discloses “*generating a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user’s future input to interjecting queries or requests for clarification if the user has the supporter role*” and “*a response builder configured to generate a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user's future input to interjecting queries or requests for clarification if the user has the supporter role.*” These claim limitations are substantively the same for purposes of my analysis so I have analyzed them together.

183. As discussed above, in the combined system of Smith & Hipp and Guinn, where the computer is determined to have more knowledge and thus has

initiative control (*i.e.*, the user is the supporter), the system operates in **directive** mode where the “computer has complete dialog control.” (EX-1005, 282.) In **directive** mode, “[u]nless the user explicitly needs some type of clarification, the computer will select its response solely according to its next goal for the task. If the user expresses need for clarification about the previous goal, this must be addressed first. No interruptions to other subdialogs are allowed.” (EX-1005, 289.) In other words, in **directive** mode, the user’s future input is limited to “*interjecting queries or requests for clarification if the user has the supporter role.*”

184. In **directive** mode, the combined system of Smith & Hipp and Guinn achieves this limit on the user’s future input by using more assertive statements (*e.g.*, commands). This can be seen in the sample dialog from Smith & Hipp below:

1. Computer: Put the knob to one zero.
2. User: Okay.
3. Computer: What is the voltage between connector one two one and connector three four?
4. User: I do not know.
5. Computer: Locate the voltmeter.
6. User: Done.
7. Computer: Add a wire between the minus com hole on the voltmeter and the connector one two one.
8. User: How do I do that?

9. Computer: Connect the end of the black wire with the small plug to the minus corn hole on the voltmeter.

10. User: Done.

11. Computer: Connect the end of the black wire with the large plug to connector one two one.

12. User: Which knob?

13. Computer: Near the top boundary and in the center is a green region. Now, in the middle right of this green region is the control knob.

14. User: Okay.

15. Computer: Put the knob to zero one.

16. User: Done.

17. Computer: Connect the end of the black wire with the large plug to connector one two one.

18. User: Where is connector three four?

19. Computer: Near the left boundary and toward the middle is a green region. Now, in the bottom center of this green region is connector three four.

20. User: Okay.

21. Computer: What is the voltage between connector one two one and connector three four?

22. User: Zero.

(EX-1005, 284-85.)

185. Accordingly, Smith & Hipp discloses “generating a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user’s future input to interjecting queries or requests for clarification if the user has the supporter role” and “a response builder configured to generate a response to the natural language utterance with a format based on the conversation type, wherein the format is adapted to limit the user’s future input to interjecting queries or requests for clarification if the user has the supporter role.”

**C. Claims 2 and 11 – “format of the response is adapted to provide discrete piece of information, ... a plurality of pieces of information, or obtain information.”**

[2A] The method recited in claim 1, wherein **determining the conversation type comprises determining a conversational goal corresponding to the natural language utterance,**

[2B] **wherein the format of the response is adapted to provide a discrete piece of information to the user, to provide a plurality of pieces of related information to the user, or to obtain one or more pieces of information from the user.**

[11A] The system recited in claim 10, wherein the speech engine is configured to **determine the conversation type further based on a conversational goal corresponding to the natural language utterance,**

[11B] **and wherein the format of the response is adapted to provide a discrete piece of information to the user, to provide a plurality of pieces of related information to the user, or to obtain one or more pieces of information from the user.**

186. Claims 2 and 11 depend from claims 1 and 11, respectively, and add two additional requirements: (a) that determining the “conversation type” includes determining a conversational goal corresponding to the user’s natural language utterance, and (b) that the format of the system’s response is adapted according to that goal. With respect to the second requirement, Claims 2 and 11 specifically provide that the response format is adapted either (i) to provide a discrete piece of information to the user, (ii) to provide a plurality of related pieces of information to the user, or (iii) to obtain one or more pieces of information from the user. In my opinion, all of these limitations of claims 2 and 11 would have been obvious to a person of ordinary skill in the art in view of the teachings of Smith & Hipp. That is because Smith & Hipp describes a goal-oriented dialogue system that inherently determines the user’s conversational goal and adapts its responses accordingly, including giving short answers, providing expanded informational answers, or asking the user for information. I will address each aspect of claims 2 and 11 in turn and map each to Smith & Hipp and Guinn’s disclosure below. The limitations of claims 2 and 11 are identical in substance, so I address them together here.

187. *Conversational goal corresponding to the utterance.* Smith & Hipp teaches that an effective dialog system must interpret each user utterance in light of the user’s goal or purpose. The Smith & Hipp system is explicitly described as “goal-oriented,” (EX-1005, 285), where efficient dialog requires that each participant

(including the computer) understand “the purpose of the interaction,” (EX-1005, 282). In other words, the system must determine the conversational goal underlying the user’s utterance in order to decide how to respond.

188. Consistent with this, Smith & Hipp describes its disclosed system as follows:

The purpose of the architecture is to deliver to users in real time the behaviors needed for efficient human-machine dialog. Specifically, it is aimed at achieving the following:

*Convergence to a goal.* Efficient dialog requires that each participant understand the purpose of the interaction and have the necessary prerequisites to cooperate in its achievement. This is the intentional structure of Grosz and Sidner (1986), the goal-oriented mechanism that gives direction to the interaction. The primary required facilities are a problem solver that can deduce the necessary action sequences and a set of subsystems capable of carrying out those sequences.

(EX-1005, 282.)

189. Smith & Hipp also describes how its “dialog controller,” which it describes as “the overall ‘boss’ of the dialog processing system,” “formulates goals at the top level to be passed on to the theorem-proving stage.” (EX-1005, 297.)

Namely, it provides:

*Dialog controller.* This is the overall “boss” of the dialog processing system. It formulates goals at the top level to be passed on to the theorem-proving stage. It determines the role that the computer plays

in the dialog by determining how user inputs relate to already established dialog as well as determining the type of response system. It also maintains all dialog information shared by the other modules and controls their activation. Its control algorithm is the highest-level dialog processing algorithms.

The basic cycle followed by the dialog controller is shown below.

- (1) Obtain suggested goal from the domain processor.
- (2) Based on the suggested goal and the current state of the dialog, select the next goal to be pursued by the computer and determine the expectations associated with that goal. (The goal may thus be selected from one of the active subdialogs. The choice is partially dependent on the current level of initiative.)
- (3) Attempt to complete the goal using the IPSIM system, possibly invoking voice interactions.
- (4) Update system knowledge based on efforts at goal completion.
- (5) Determine next operations to be performed by the domain processor in providing a suggested goal.
- (6) Go to step 1.

(EX-1005, 297.)

190. Smith & Hipp also discusses prior literature as follows:

Whittaker and Stenton (1988) propose a definition of **dialog control based on the utterance type of the speaker: question, assertion, command, prompt**. In the first three, the speaker usually has control

unless the utterance is a direct response to a question. In the last, control is usually being relinquished. They note that there is a close relationship between topic shift and control shift, and thus dialog control is useful in identifying the discourse structure.

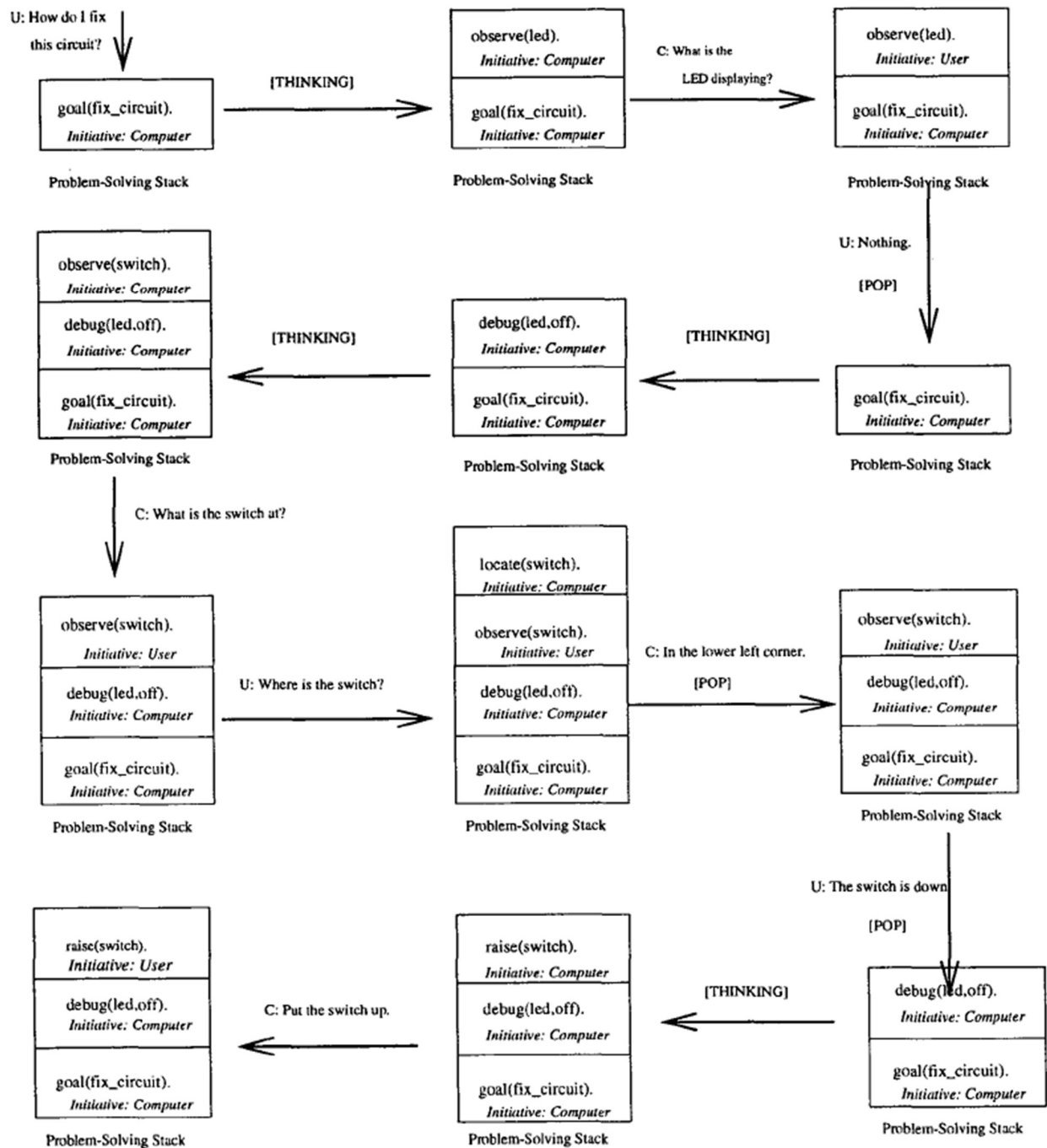
Kitano and Van Ess-Dykema (1991) extend the plan recognition model of Litman and Allen (1987) to consider variable initiative dialog. **In addition, they extend the initiative control rules of Whittaker and Stenton to consider utterance content.** They note that when a speaker makes an utterance that is relevant to **his or her speaker-specific domain plan, then that speaker has dialog control.** Their observation that **there are speaker-specific plans or goals is crucial to our proposed model for participating in variable initiative dialogs.** Selecting the next subdialog to be entered may require selecting between differing computer and user goals. This selection process must be a function of the current level of dialog initiative.

(EX-1005, 315.) In other words, Smith & Hipp recognized that variable initiative modes could be determined depending on speaker-specific plans or goals, just like the conversation types of the '765 patent.

191. Guinn likewise directly supports the step of determining a conversational goal corresponding to the natural language utterance by tying utterance interpretation to a structured task model. Guinn explains that “initiative levels for each goal are defined during the interaction based on 1) explicit and implicit initiative-changing utterances and 2) competency evaluation,” and that “an

agent may have initiative over one goal but not another. As goals get pushed and popped from the problem-solving stack, initiative changes accordingly.” (EX-1006, 279.) The end result of this, per Guinn, is that “many initiative changes are done implicitly based on which goal is being solved.” (EX-1006, 279.) A POSITA would thus understand from Guinn that in order to determine initiative for a conversation (*i.e.*, the conversation type), the system must first understand the conversational goal.

192. Guinn illustrates this with dialogue examples where utterances like “How do I fix this circuit?” trigger the goal “goal(fix\_circuit),” and subsequent user responses lead to goal-specific subdialogues such as “observe(switch)” or “debug(led.off).” (EX-1006, 280.) This is illustrated below:



Guinn, Figure 1 (280)

193. Guinn further explains how this figure shows that “all initiative changes can be accounted for by explicit initiative-changing utterances or by popping of the problem-solving stack due to goal resolution.” (EX-1006, 279). This

goal-based interpretation of user input demonstrates that Guinn's system actively determines conversational goals from natural language, just as required by the limitation.

194. Smith & Hipp teaches response format adaptation. Specifically, Smith & Hipp discloses:

The system must output statements that are compatible with its level of initiative. This means that the output generator must have a parameter that enables the system to specify assertiveness. Examples of the request to "turn the switch up" at various levels of assertiveness are as follows:

|                                        |                     |
|----------------------------------------|---------------------|
| Turn the switch up.                    | (command)           |
| Would you [please] turn the switch up? | (request)           |
| Can you turn the switch up?            | (indirect request)  |
| The switch can be turned up.           | (statement of fact) |
| Turning the switch up is necessary.    | (statement of fact) |

Two examples of querying for the switch position are as follows:

|                                     |                    |
|-------------------------------------|--------------------|
| What is the switch position?        | (request)          |
| I need to know the switch position. | (indirect request) |

(EX-1005, 289-290.)

195. Thus, Smith & Hipp teaches a response format adaptation based on four levels of assertiveness as command, request, indirect request, and statement of

fact. A POSITA would understand that Smith & Hipp teaches that the system generates responses that can provide discrete pieces of information in the “statement of fact” assertive level, or obtain information from the user through “request” or “indirect request” assertive levels.

196. *Response format adapted to provide a discrete piece of information.* Smith & Hipp also teaches adjusting the format and content of the system’s response depending on the user’s needs and the dialog context. In one scenario, the system provides a single, discrete piece of information in response to a user question – *i.e.* a concise answer. For instance, Smith & Hipp describes a “Passive Mode” of interaction in which “[the] computer will ... provide information only as a direct response to a user question.” (EX-1005, 289.) Put another way, in this mode, the system confines its response to the specific information asked for, without volunteering additional details. (EX-1005, 289.) A POSITA would understand that this often corresponds to the format of response that “*provides a discrete piece of information to the user,*” as recited in claims 2 and 11.

197. Smith & Hipp also includes examples of disclosures meeting this limitation. For instance, Smith & Hipp provides the following sample dialog of the system operating in declarative mode:<sup>2</sup>

---

<sup>2</sup> Depending on the context of the conversation, the system may also provide short and direct responses in other initiative modes as well, particularly where the

1. Computer: This is the circuit-fixit-shop. How may I help you?
2. User: RS111 is not working.
3. Computer: I am familiar with that circuit. The LED is supposed to be displaying alternately flashing one and seven.
4. User: There is no wire on one zero four.
5. Computer: There is supposed to be a wire between connector one zero four and connector one zero two.
6. User: The circuit is working.
7. Computer: Glad to have been of assistance. Good-bye.

(EX-1005, 285.) Here, the system is configured to provide a “*discrete piece of information to the user,*” namely that there is a wire between two specific connectors.

198. *Response format adapted to provide a plurality of pieces of related information.* Smith & Hipp also describes a mode in which the system’s response can be more expansive, providing multiple pieces of related information when appropriate. Namely, in the **declarative** mode, the computer is “free to mention relevant, though not required, facts as a response to the user’s statements.” (EX-1005, 289.) In other words, the system can go beyond a simple answer and include

---

user is in control of the dialog (*e.g.*, in declarative mode). In other words, the Smith & Hipp system does not only “*provide a discrete piece of information to the user*” in passive mode.

additional related facts in its reply. This is exactly a response format “*adapted ... to provide a plurality of pieces of related information to the user.*”

199. Additionally, in Smith & Hipp’s directive mode, the computer is in complete control and “will use whatever dialog is necessary to obtain the needed item of knowledge related to the subgoal.” (EX-1005, 282.) Thus, in directive mode, the computer can provide multiple pieces of information to the user for the user to assist in achieving the goal. This response format is also “*adapted ... to provide a plurality of pieces of related information to the user.*”

200. Smith & Hipp provides examples of disclosures meeting this limitation as well. For instance, it provides the following sample dialog of the system operating in directive mode:

1. Computer: Put the knob to one zero.
2. User: Okay.
3. Computer: What is the voltage between connector one two one and connector three four?
4. User: I do not know.
5. Computer: Locate the voltmeter.
6. User: Done.
7. Computer: Add a wire between the minus com hole on the voltmeter and the connector one two one.
8. User: How do I do that?

9. Computer: Connect the end of the black wire with the small plug to the minus corn hole on the voltmeter.

10. User: Done.

11. Computer: Connect the end of the black wire with the large plug to connector one two one.

12. User: Which knob?

13. Computer: Near the top boundary and in the center is a green region. Now, in the middle right of this green region is the control knob.

14. User: Okay.

15. Computer: Put the knob to zero one.

16. User: Done.

17. Computer: Connect the end of the black wire with the large plug to connector one two one.

18. User: Where is connector three four?

19. Computer: Near the left boundary and toward the middle is a green region. Now, in the bottom center of this green region is connector three four.

20. User: Okay.

21. Computer: What is the voltage between connector one two one and connector three four?

22. User: Zero.

(EX-1005, 284-85.)

201. In each of the underlined responses above, the system provides multiple pieces of related information in response to the user. Thus, the response format is *“adapted ... to provide a plurality of pieces of related information to the user.”*

202. *Response format adapted to obtain information from the user.* Smith & Hipp further discloses that its dialog system can adapt its response in a manner suitable *“to obtain one or more pieces of information from the user.”*

203. Indeed, a critical aspect of the dialog architecture disclosed by Smith & Hipp is the ability of the system to discern whether it needs additional data to achieve a sub-goal. For example, in the described “Directive Mode” of operation, the computer takes the lead in the dialog and “will use whatever dialog is necessary to obtain the needed item of knowledge related to the subgoal.” (EX-1005, 282.) In practice, this means the system can ask the user one or more questions to gather missing information. For example, here the system is asking the user to provide missing pieces of information (the voltage between two connectors):

1. Computer: Put the knob to one zero.
2. User: Okay.
3. Computer: What is the voltage between connector one two one and connector three four?
4. User: I do not know.
5. Computer: Locate the voltmeter.
6. User: Done.

7. Computer: Add a wire between the minus com hole on the voltmeter and the connector one two one.
8. User: How do I do that?
9. Computer: Connect the end of the black wire with the small plug to the minus corn hole on the voltmeter.
10. User: Done.
11. Computer: Connect the end of the black wire with the large plug to connector one two one.
12. User: Which knob?
13. Computer: Near the top boundary and in the center is a green region. Now, in the middle right of this green region is the control knob.
14. User: Okay.
15. Computer: Put the knob to zero one.
16. User: Done.
17. Computer: Connect the end of the black wire with the large plug to connector one two one.
18. User: Where is connector three four?
19. Computer: Near the left boundary and toward the middle is a green region. Now, in the bottom center of this green region is connector three four.
20. User: Okay.
21. Computer: What is the voltage between connector one two one and connector three four?

22. User: Zero.

(EX-1005, 284-85.)

204. Likewise, Smith & Hipp further describes a “missing axiom” technique where, if the automated reasoning requires a piece of information it lacks, the system engages in a dialog to acquire that information from the user. (EX-1005, 286.) As Smith & Hipp explains:

The third possibility is that neither find(swl) or reportposition(swl,-) would exist in the database, in which case both could be sent to the controller as missing axioms for possible vocalization. (The decision as to whether to send a missing axiom to the controller depends on whether the axiom represents an answerable question by the user. The system maintains a mechanism for indicating what is reasonable to ask and what is not, as described below.)

(EX-1005, 286.)

205. Thus Smith & Hipp discloses “*determining a conversational goal corresponding to the natural language utterance, wherein the format of the response is adapted to provide a discrete piece of information to the user, to provide a plurality of pieces of related information to the user, or to obtain one or more pieces of information from the user.*”

**D. Claim 3 – “discrete” responses shorter than “plurality of pieces” responses**

3. The method recited in claim 2, wherein the response generated in the format that is adapted to provide the discrete piece of information is shorter than the response generated in the format that is adapted to provide the plurality of pieces of related information.

206. Claim 3 depends from claim 2 and further recites “*wherein the response generated in the format that is adapted to provide the discrete piece of information is shorter than the response generated in the format that is adapted to provide the plurality of pieces of related information.*”

207. In simple terms, claim 3 is saying that a brief answer providing one fact should be shorter in length than a more detailed answer providing many facts. This is a logical expectation and a natural result of the claim 2 features, and Smith & Hipp’s teachings make this relationship obvious as well.

208. Smith & Hipp implies that a one-piece informational response will indeed be shorter and more succinct than a response containing multiple pieces of information. As discussed above, in Smith & Hipp’s system, a direct answer to a user’s question (the discrete piece of information format), such as would be provided under the **passive** mode, involves the computer giving *only* the information requested *and nothing more*. (EX-1005, 283.) Such a response is typically a single sentence or a brief statement confined to the specific answer. In contrast, when the system provides additional related facts (*i.e.*, “the plurality of related information”

format), the response naturally becomes longer, often comprising multiple sentences or a more elaborate explanation to convey those extra details. For example, a concise answer like “The switch is on the front panel” contains one piece of information and would be much shorter than an expanded response that provides several related facts (e.g., “The switch is on the front panel, just above the indicator light, and it controls power to the unit”), which contains multiple pieces of information.

209. Smith & Hipp also provides examples consistent with this limitation. Indeed, consider once again the examples provided above in the context of claims 2 and 11. Here is the example I provided previously of a response “adapted to provide a discrete piece of information to the user”:

1. Computer: This is the circuit-fixit-shop. How may I help you?
2. User: RSIII is not working.
3. Computer: I am familiar with that circuit. The LED is supposed to be displaying alternately flashing one and seven.

(EX-1005, 285.) And here is the example I provided previously of a response “adapted ... to provide a plurality of pieces of related information to the user”:

11. Computer: Connect the end of the black wire with the large plug to connector one two one.
12. User: Which knob?
13. Computer: Near the top boundary and in the center is a green region. Now, in the middle right of this green region is the control knob.

14. User: Okay.
15. Computer: Put the knob to zero one.
16. User: Done.
17. Computer: Connect the end of the black wire with the large plug to connector one two one.
18. User: Where is connector three four?
19. Computer: Near the left boundary and toward the middle is a green region. Now, in the bottom center of this green region is connector three four.
20. User: Okay.
21. Computer: What is the voltage between connector one two one and connector three four?
22. User: Zero.

(EX-1005, 284-85.)

210. As is evident, the computer's provision of a short, discrete piece of information ("The LED is supposed to be displaying alternately flashing one and seven" (12 words)) is "shorter" than the provision of a plurality of a pieces of information ("Near the top boundary and in the center is a green region. Now, in the middle right of this green region is the control knob" or "Near the left boundary and toward the middle is a green region. Now, in the bottom center of this green region is connector three four." (each 25 words)).

211. Thus, the combination of Smith & Hipp and Guinn discloses “a response generated in the format that is adapted to provide the discrete piece of information is shorter than the response generated in the format that is adapted to provide the plurality of pieces of related information.”

**E. Claims 4 and 12 – “determining conversation type” based on “conversational goal.”**

[4A] The method recited in claim 1, wherein **determining the conversation type includes determining a conversational goal corresponding to the natural language utterance,**

[4B] **wherein the conversation type is determined as a query conversation if the conversational goal is to gather discrete information or perform a discrete task, is determined as a didactic conversation if the conversational goal is to gather related information, or is determined as an exploratory conversation if the natural language utterance has no specific goal.**

[12A] The system recited in claim 10, wherein the speech engine is configured to **determine the conversation type further based on a conversational goal corresponding to the natural language utterance,**

[12B] **and wherein the conversation type is determined to be a query conversation if the conversational goal is to gather discrete information or perform a discrete task, is determined to be a didactic conversation if the conversational goal is to gather related information, or is determined to be an exploratory conversation if the natural language utterance has no specific goal.**

212. Claims 4 and 12 depend from claims 1 and 10 and further recite: “wherein determining the conversation type includes determining a conversational goal corresponding to the natural language utterance, wherein the conversation

*type is determined as a query conversation if the conversational goal is to gather discrete information or perform a discrete task, is determined as a didactic conversation if the conversational goal is to gather related information, or is determined as an exploratory conversation if the natural language utterance has no specific goal.”*

213. In simple terms, claims 4 and 12 are saying that (1) a conversation type is determined with the benefit of a determination of a conversational goal, and (2) more specifically, if the user just wants a quick fact or to do one simple thing, it’s a “query” conversation; if the user is trying to understand a process or learn something through a series of connected questions, it’s a “didactic” conversation; and if the user is just exploring or talking without a clear goal, it’s an “exploratory” conversation.

214. I have previously discussed in the context of claims 2 and 11 how the Smith & Hipp and Guinn combination discloses or at least renders obvious *“determining the conversation type comprises determining a conversational goal corresponding to the natural language utterance.”* That language is identical to the language here in claims 4 and 12 and I therefore incorporate my prior analysis by reference.

215. However, I note here additional disclosures that are relevant both to claims 4 and 12, but claims 2 and 11 as well. Namely, as Smith and Hipp acknowledge, there had been extensive research into the relationship between

conversational goals and dialog control for years even prior to the Smith and Hippi publication. For example, as Smith and Hippi explain:

Whittaker and Stenton (1988) propose a definition of dialog control based on the utterance type of the speaker: question, assertion, command, prompt. In the first three, the speaker usually has control unless the utterance is a direct response to a question. In the last, control is usually being relinquished. They note that there is a close relationship between topic shift and control shift, and thus dialog control is useful in identifying the discourse structure.

Kitano and Van Ess-Dykema (1991) extend the plan recognition model of Litman and Allen (1987) to consider variable initiative dialog. In addition, they extend the initiative control rules of Whittaker and Stenton to consider utterance content. They note that when a speaker makes an utterance that is relevant to his or her speaker-specific domain plan, then that speaker has dialog control. Their observation that there are speaker-specific plans or goals is crucial to our proposed model for participating in variable initiative dialogs. Selecting the next subdialog to be entered may require selecting between differing computer and user goals. This selection process must be a function of the current level of dialog initiative.”

(EX-1005, 315.) Put differently, variable initiative modes can be determined based on speaker-specific plans or goals.

216. Further, as Smith and Hippi also disclose even older research by Grosz and Sidner that relates to the idea of intentionality in dialog:

The intentional component specifies the purpose of the dialog. Grosz and Sidner (1986) use the notation DP and DSP to stand for "discourse purpose" and "discourse segment purpose." These entities correspond to the predicate goals that our system poses and then builds the dialog around. Our system, in fact, constructs a set of partial proofs and these are our instantiation of the intentional component.

(EX-1005, 282.) This reveals the connection between the so-called “predicate goals” (*i.e.*, conversational goals) and that the Smith and Hipp system “builds the dialog around.”

217. I turn now to the second aspect of the claims: linking specific conversational goals to specific conversation types.

218. “*Query*” if goal is to gather discrete information or perform a discrete task. Claim 4 specifies that if the conversational goal is to gather discrete information or perform a discrete task, the conversation type is determined as a “*query conversation*.” Smith & Hipp discloses this limitation in its description of its **passive** mode. As discussed above, Smith & Hipp’s passive mode corresponds to the query conversation type of the ’765 patent. Smith & Hipp describes the **passive** mode as reflecting the instance where “[t]he computer responds directly to user questions ... without recommending a subgoal as the next course of action.” (EX-1005, 283.) This disclosure directly corresponds with this element of claims 4 and 12.

219. *Didactic if goal is to gather related information.* Claims 4 and 12 also provide that if the conversational goal is to gather related information, the conversation type is determined as a “*didactic conversation.*” In other words, when the user’s aim is to obtain a broader set of information on a topic (rather than the single fact requested by the user), the interaction is classified as didactic. As noted above, Smith & Hipp’s directive mode corresponds to the didactic conversation type of the ’765 patent. In directive mode, the computer “will use whatever dialog is necessary to obtain the needed item of knowledge related to the subgoal.” (EX-1005, 282.) The conversational goal in directive mode is therefore “*to gather related information.*”

220. *Exploratory if no specific goal.* Finally, claims 4 and 12 state that if the natural language utterance has no specific goal, the conversation type is determined as an “*exploratory conversation.*” As noted above, Smith & Hipp’s suggestive and declarative modes correspond to the exploratory conversation of the ’765 patent, with the difference between the two modes being whether the control bias is in the user’s favor (declarative) or the computer’s favor (suggestive). (§IV.B.3.) In both of these modes, there is no specific goal set, just like an exploratory conversation. For example, in suggestive mode, “the computer will make suggestions about the subgoal to perform next, but it is also willing to change the direction of the dialog according to stated user preferences.” (EX-1005, 283.)

221. Thus the combination of Smith & Hipp and Guinn discloses “determining the conversation type includes determining a conversational goal corresponding to the natural language utterance, wherein the conversation type is determined as a query conversation if the conversational goal is to gather discrete information or perform a discrete task, is determined as a didactic conversation if the conversational goal is to gather related information, or is determined as an exploratory conversation if the natural language utterance has no specific goal.”

**F. Claims 5 and 13 – “conversation type” determined based on “role”**

5. The method recited in claim 1, wherein the conversation type is determined as a query conversation if the user that spoke the natural language utterance has the leader role, is determined as a didactic conversation if the user that spoke the natural language utterance has the supporter role, or is determined as an exploratory conversation if the user that spoke the natural language utterance has both the leader role and the supporter role.

13. The system recited in claim 10, wherein the speech engine is configured to determine the conversation type based on the role of the user, wherein the conversation type is determined to be a query conversation if the user that spoke the natural language utterance has the leader role, is determined to be a didactic conversation if the user that spoke the natural language utterance has the supporter role, or is determined to be an exploratory conversation if the user that spoke the natural language utterance has both the leader role and the supporter role.

222. The limitations of claims 5 and 13 are identical in substance, so I address them together here.

223. In simple terms, claims 5 and 13, which depend from claims 1 and 10, recite that the type of conversation depends on the role the user is playing. If the user is acting as the leader—driving the dialog and asking for specific facts or actions—it’s a “query” conversation. If the user is acting more like a supporter—helping the system reach its goal or supplying related information—it’s a “didactic” conversation. And if the user is both leading and supporting at the same time, then the conversation is considered “exploratory”. This distinction reflects how initiative and control shift in dialog and tracks the disclosures made in Smith & Hipp.

224. *Query if User Leads.* Claims 5 and 13 specify that if the user has the leader role, the conversation type is a query conversation. Smith & Hipp teaches that when the user controls the dialogue – *i.e.* the user is taking the lead – the system operates in a completely **passive** mode, essentially following the user’s queries and instructions. (EX-1005, 282-83.) As Smith & Hipp describes, in the **passive** mode, “[t]he user has complete dialog control. The computer responds directly to user questions and passively acknowledges user statements without recommending a subgoal as the next course of action.” (EX-1005, 283.) A POSITA would immediately understand that in Smith & Hipp’s **passive** mode, the user is the “leader” of the conversation, as that term is described in the ’765 patent.

225. *Didactic if User Supports.* Claims 5 and 13 next state that if the user has the supporter role, the conversation type is a didactic conversation. In this case,

the system is the one leading the conversation, and the user is primarily listening or responding as needed – effectively being guided by the system. Smith & Hipp likewise describes such situations. In its highest system-initiative mode (described as “directive”), “**the computer has complete dialog control**” and directs the course of the conversation to achieve a subgoal. (EX-1005, 282.) The user in that mode is essentially a supporter, following the system’s prompts or instructions. The nature of this interaction is didactic: the system is imparting information or instructing the user to accomplish tasks, much like a teacher or guide. A POSITA would equate Smith & Hipp’s system-led dialog mode with a didactic conversation, since the goal in such mode is to convey information to the less-informed party (the user) in a controlled manner. Therefore, when the user occupies the supporter role (and the system leads), Smith & Hipp makes it obvious that the conversation is of a **directive** (or in the lingo of the ’765 patent, didactic) type. (EX-1005, 282.)

226. *Exploratory if user both leads and supports.* Finally, claims 5 and 13 provide that if the user has both the leader and the supporter role, the conversation type is an exploratory conversation. This limitation is obvious based on Smith & Hipp’s model of variable initiative dialog. The system explicitly defines modes—**suggestive** and **declarative**—in which the user and the system may alternately assume initiative over dialogs, depending on the user’s knowledge and behavior.

227. In **declarative** mode, for example, “the user has dialog control, but the computer is free to mention relevant, though not required, facts.” (EX-1005, 283). In this configuration, the user initially leads the conversation—introducing new subgoals and queries—but also can take a supporter role to the system by answering its questions and complying with task steps. Smith & Hipp notes this explicitly: “The user may ignore suggestions by following his or her own preferred dialog paths,” but the system “tracks the selected topic and responds in an appropriate manner.” (EX-1005, 286.) Thus, the user alternates between filling the roles of leader (goal initiator) and supporter (goal executor)—precisely as described in the '765 limitation. The sample dialogs illustrated on pages 284-286 show how this back-and-forth can work.

228. Guinn further reinforces this concept through its model of initiative and task control. It states that “dialogue initiative should always pass immediately to the participant who is best able to handle the current task,” (EX-1006, 278), which creates a dynamic where the user may lead on some goals and support the system on others, even within the same utterance sequence. Guinn formalizes this using explicit and implicit initiative-changing utterances, in which users can take control by asking questions or introducing new goals, (EX-1006, 279), while also reverting to a supportive role when the system’s knowledge exceeds their own.

229. Guinn’s description of explicit initiative changing based on who is asking questions explains this further. Guinn states: “Our model concentrates on two specific dialogue cues: questions and answers. When an agent  $A_1$  asks another agent  $A_2$  to satisfy a goal  $G$ , agent  $A_2$  gains initiative over goal  $G$  and all subgoals of  $G$  until agent  $A_2$  passes control of one of those subgoals back to agent  $A_1$ .” (EX-1006, 279.) A POSITA would understand from Guinn and Smith & Hipp that an initiative mode that alternates between purely system-led and purely user-led would reflect a scenario where, in the course of a single conversation, a user is taking on both a leader and supporter role.

230. Guinn’s initiative-setting mode explicitly models this behavior: control shifts back and forth depending on relative task competency. (*See* EX-1006, 281.) This mirrors the condition in Claim 6, where the user plays both a leading and assisting role during a conversation. By integrating goal-based initiative shifts with user competency models, both Smith & Hipp and Guinn teach a framework where the user naturally inhabits both the leader and supporter roles in mixed-initiative, exploratory dialogues. Therefore, a POSITA would find it obvious over a combination of Smith & Hipp and Guinn that a system recognizing this dynamic would classify the dialog as exploratory under the conditions described.

**G. Claims 6, 7, and 14 – “conversation type” determined based on “allocation of information.”**

6. The method recited in claim 1, wherein determining the conversation type is based on an allocation of information between the user and the speech engine, wherein the conversation type is determined to be a query conversation or a didactic conversation if the user that spoke the natural language utterance has a lower allocation of information related to the natural language utterance than the speech engine, or is determined as a didactic conversation if the user and the speech engine have an equal or alternating allocation of information related to the natural language utterance.

7. The method recited in claim 1, wherein determining the conversation type further comprises determining an allocation of information between the user and the speech engine.

14. The system recited in claim 10, wherein the speech engine is configured to determine the conversation type based on an allocation of information between the user and the speech engine, wherein the conversation type is determined to be a query conversation or a didactic conversation if the user that spoke the natural language utterance has a lower allocation of information related to the natural language utterance than the speech engine or is determined to be a didactic conversation if the user and the speech engine have an equal or alternating allocation of information.

231. Claims 6, 7, and 14 recite determining the conversation type is based on an allocation of information between the user and the speech engine and claims 6 and 14 also provide a mapping between the allocation of information and a specific conversation type.

232. Based on my review of the '765 patent and the intrinsic evidence, if claims 6 and 14 do not include a drafting error then they do not make sense. For

example, the claims describes a “didactic” conversation in the way that the specification describes an “exploratory” one. If there is a drafting error, then claims 6 and 14 should read: “is determined as a ~~didactic~~ **exploratory** conversation if the user and the speech engine have an equal or alternating allocation of information related to the natural language utterance.”

233. If there is a drafting error, in simple terms, claims 6, 7, and 14, which depend from claims 1 and 10, are reciting that (1) a conversation type is determined with the benefit of a determination of an allocation of information between the user and the speech engine, and, (in the case of claims 6 and 14 only) (2) if the system knows more than the user, the conversation is classified as either a query or didactic conversation type, but if the user and system share the knowledge equally or take turns contributing information, then it’s an “exploratory” conversation. I address these two components in turn.

234. Starting with using allocation of information to determine the conversation type, Smith & Hipp emphasizes accounting for the distribution of knowledge between the user and the system throughout a dialogue. As described, the system maintains a dynamic user model that tracks the information the user knows or does not know in order to keep an “accounting for user knowledge and abilities.” (EX-1005, 281-82.) This allows the system to gauge the user’s knowledge state for the current topic or goal. A POSITA would appreciate that Smith & Hipp’s

approach inherently teaches considering which party has more information about the subject of the utterance when deciding how the dialogue should proceed – directly paralleling the concept in claim 6 of determining conversation type based on the user’s vs. system’s information allocation.

235. Smith & Hipp further discloses maintaining a dynamic “user model” that keeps track of information known to the user during a dialogue. (EX-1005, 282, 287-88.) The user model’s purpose is “to indicate what needs to be said to the user to enable the user to function effectively. It also indicates what should be omitted because of existing user knowledge.” (EX-1005, 282.) In other words, Smith & Hipp’s system continually assesses which pieces of information (*i.e.*, knowledge) are held by the user versus by the system. As the dialogue progresses, “[m]entioned facts are stored in the [user] model as known to the user and are not repeated. Previously unmentioned information may be assumed to be unknown and [is] explained as needed.” (EX-1005, 282.)

236. Smith & Hipp further describes how user questions “may indicate lack of knowledge and result in the removal of items from the user model.” (EX-1005, 282.) By tracking what the user knows and does not know, Smith & Hipp determines an allocation of information between the user and the system (speech engine) at each step of the conversation.

237. Smith & Hipp further discloses that the allocation of information between the user and system can be directly relevant to determining who should control the conversation—whether the system should lead or the user should. In Smith & Hipp’s architecture, there should be a “change of initiative” based on which party holds the requisite knowledge to advance the task at hand. (EX-1005, 282.)

238. As Smith & Hipp further explains: sometimes “an efficient interaction requires that the machine yield control so that the more competent partner can lead the way to the fastest possible solution.” (EX-1005, 282.) This determination relies on the system’s user model, which “specifies information needed for efficient interaction” and “indicates what needs to be said to the user ... [and] what should be omitted because of existing user knowledge.” (EX-1005, 282.) If the system recognizes that the user has sufficient knowledge, it may “yield to the user’s control and respond cooperatively to his or her requests.” (EX-1005, 282.)

239. Conversely, if the user lacks information, the system “must be able to carry out its own problem-solving process and direct the actions to task completion.” (EX-1005, 282.) Thus, by continually evaluating and updating the allocation of information between the participants, the system determines whether to maintain control or relinquish it, ensuring more efficient, context-sensitive dialog management.

240. A POSITA would recognize that using such a user model to inform dialog management is equivalent to determining the allocation of information between the user and the speech engine as required by claims 6, 7, and 14. Indeed, Smith & Hipp emphasizes that accounting for the user's knowledge is "needed for efficient interaction" in a voice dialog system, (EX-1005, 282), which would motivate a POSITA to include this feature when determining a conversation's type or strategy.

241. While the Smith & Hipp system, like the '765 patent, did not disclose an automatic process by which conversation type was updated mid-conversation to reflect an allocation of information, Guinn provides that disclosure if it is needed. Namely, as I previously explained (§IV.A.2), Guinn discloses multiple suitable algorithms for assigning initiative modes (*i.e.*, conversation types) based on which participant is the "more knowledgeable agent" as defined by, for example, "which agent has the greater total percentage of knowledge." (EX-1006, 281.) Thus the Smith & Hipp-Guinn combination fully discloses the limitation in claims 6, 7, and 14 of "determining the conversation type based on an allocation of information between the user and the speech engine."

242. Smith & Hipp also discloses the specific mapping between the allocation of information and the three conversation types disclosed in claims 6 and 14: query, didactic, and exploratory. For the ease of the reader I map out that

mapping below (assuming that there is in fact the drafting error I highlighted previously):

| Condition                                                                                                                                                      | Conversation Type |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| “if the user that spoke the natural language utterance has a lower allocation of information related to the natural language utterance than the speech engine” | Query or Didactic |
| “if the user and the speech engine have an equal or alternating allocation of information related to the natural language utterance.”                          | Exploratory       |

243. As an initial matter, I note that I have been informed by counsel that, where a claim contains conditional language, only one of the recited conditions needs to be found in the prior art for the claim to be unpatentable. Thus, for example, I understand that, for the prior art to render obvious claim 6, the prior art need only satisfy one of the two identified conditions to render obvious the claim, *i.e.*, the conversation type “*is determined to be a query conversation or a didactic conversation if the user that spoke the natural language utterance has a lower allocation of information related to the natural language utterance than the speech engine*” **or** “*is determined as a [exploratory] conversation if the user and the speech engine have an equal or alternating allocation of information related to the natural language utterance.*” However, as explained below, in all events both conditions are rendered obvious over Smith & Hipp.

244. *User less knowledgeable means query or didactic.* Claim 6 first specifies that if the user has a lower allocation of information related to the utterance

than the speech engine, the conversation type is determined to be a query conversation or a didactic conversation. In other words, when the system knows more about the topic than the user does, the interaction will be classified as either query-type or didactic. Smith & Hipp plainly teaches this scenario. If the user is relatively uninformed (lacking certain knowledge) and the system holds that information, two outcomes are possible: the user can ask the system for the information, or the system can tell/teach the user the information. Smith & Hipp accommodates both cases. It notes that the system will provide information in response to user questions (**query** / **passive** scenario), and also that the system can take initiative to suggest or explain information needed to reach the goal (didactic/directive scenario). (EX-1005, 282-83.) A POSITA would therefore find it obvious that when the user has less information than the system on a subject, the dialogue naturally falls into either a query pattern (user questioning a knowledgeable system) or a didactic pattern (system instructing an uninformed user). Smith & Hipp supplies the reasoning and examples to arrive at this conclusion, meeting the claim 6 requirement for the case where the user's information is lower. (EX-1005, 282.)

245. Smith & Hipp further discloses examples of the speech engine determining the conversation type as, for example, **declarative** where the user has a lower allocation of information than the speech engine:

1. Computer: This is the circuit-fixit-shop. How may I help you?

2. User: RSIII is not working.
3. Computer: I am familiar with that circuit. The LED is supposed to be displaying alternately flashing one and seven.
4. User: There is no wire on one zero four.
5. Computer: There is supposed to be a wire between connector one zero four and connector one zero two.
6. User: The circuit is working.
7. Computer: Glad to have been of assistance. Good-bye.

(EX-1005, 285.) As Smith & Hipp explains: “Here the machine diagnoses as well as it can the current subdialog from user comments. Then it presents information it surmises may be helpful, specifically facts.” (EX-1005, 285.)

246. *Equal or alternating allocation of knowledge means exploratory.* To the extent the Board determines that both conditional limitations of claim 6 need to be satisfied to demonstrate obviousness, it is my opinion that the last limitation of claim 6 also is disclosed by the combined system of Smith & Hipp and Guinn based on Patent Owner’s understanding of the limitation in its infringement allegation.

247. As Smith & Hipp explains, dialog control is directly related to allocation of knowledge. Specifically, because “an efficient interaction requires that the machine yield control so that the more competent partner can lead the way to the fastest possible solution,” a POSITA would understand that the Smith & Hipp

system directly connects dialog initiative with the role of the relative information provider. (EX-1005, 282.)

248. This connection between dialog control and allocation of information is emphasized by Guinn. Indeed, Guinn discloses how allocation of knowledge should in some instances directly control dialog control. For instance:

In SingleSelection mode, the more knowledgeable agent (defined by which agent has the greater total percentage of knowledge) is given initiative.

...

In Continuous mode, the more knowledgeable agent (defined by which agent's first-ranked branch is more likely to succeed) is initially given initiative.

(EX-1006, 281.)

249. Smith & Hipp discloses an architecture that explicitly provides allows for alternating dialog control based on who has more relevant information at a given point in time. In particular, Smith & Hipp's **suggestive** and **declarative** initiative modes each provide for dialog control to alternate between user and computer as needed during a task based on a dynamic assessment of which participant has a more favorable allocation of relevant information. For example, in the **suggestive** mode, the computer "still has dialog control, but not as strongly" and in **declarative** mode,

the user “has dialog control, but the computer is free to mention relevant, though not required, facts as a response to the user's statements.” (EX-1005, 283.)

250. These configurations inherently involve an equal or alternating allocation of information—that is, both participants (user and system) introduce and respond to content. The system is not simply reacting (as in a **passive** mode) or commanding (as in **directive** mode), but dynamically engaging in an interactive, fluid conversation shaped by the conversational goal and a dynamic assessment of the allocation of information. This falls squarely within what the '765 patent refers to as “*exploratory conversation*.”

251. Again, Guinn makes this even more explicit. Guinn explains “efficient human-computer dialogue requires immediate utterance-by-utterance accommodation to the needs of the interaction. ... This means that the dialogue initiative should always pass immediately to the participant who is best able to handle the current task.” (EX-1006, 278.)

**H. Claim 8 - “only provides information requested ... limits the user’s future input”**

8. The method recited in claim 1, wherein generating the response comprises generating, **if the conversation type is a query conversation, a response that only provides information requested by the natural language utterance,** and **generating, if the conversation type is a didactic conversation, a response that controls the conversation and limits the user’s future input to a request or query for clarification.**

252. A combination of Smith & Hipp and Guinn discloses a method of generating the response that comprises “*generating, if the conversation type is a query conversation, a response that only provides information requested by the natural language utterance, and generating, if the conversation type is a didactic conversation, a response that controls the conversation and limits the user’s future input to a request or query for clarification.*”

253. Put into simpler terms, claim 8 requires the system to tailor its reply based on the conversation type: in a **query** conversation (*i.e.*, where the user leads the interaction), the system’s response should only include the information the user asked for; whereas in a **didactic** conversation (*i.e.*, where the system leads and the user is in a supporting role), the system’s response should take control of the dialogue, allowing the user to interject only with clarification questions. Smith & Hipp provides these features.

254. As discussed previously, Smith & Hipp identifies a continuum of dialog initiative modes, including at the two extremes a user-led mode (called **passive** mode) and a system-led mode (called “Directive Mode”). In the user-led **passive** mode, “the user has complete dialog control” and the computer behaves accordingly: “It will provide information only as a direct response to a user question.” (EX-1005, 289.) Smith & Hipp thus discloses that when the user is in control of the conversation (*i.e.*, in a query-type interaction), the system’s response only provides the

information requested by the user's utterance, with no extra or unsolicited information. This is exactly the behavior claim 8 requires for a "*query conversation*," *i.e.*, generating "*a response that only provides information requested by the natural language utterance.*"

255. Conversely, Smith & Hipp's **directive** mode corresponds to a system-led scenario. In **directive** mode, "[t]he computer has complete dialog control" and will drive the conversation by pursuing its own task goals, "select[ing] its response solely according to its next goal for the task," unless the user needs clarification. (EX-1005, 289.) The system does not allow the user to deviate the dialog except "[i]f the user expresses need for clarification." (EX-1005, 289.)

256. In other words, when operating in this mode, the system's responses control the course of the conversation, and the user's input is limited to clarification requests—precisely matching the didactic conversation case of claim 8, *i.e.*, generating "*a response that controls the conversation and limits the user's future input to a request or query for clarification.*"

257. Thus, a combination of Smith & Hipp and Guinn discloses "*wherein generating the response comprises generating, if the conversation type is a query conversation, a response that only provides information requested by the natural language utterance, and generating, if the conversation type is a didactic*

*conversation, a response that controls the conversation and limits the user's future input to a request or query for clarification.”*

**I. Claim 9 - “user controls the conversation or ... provides information”**

9. The method recited in claim 1, wherein the natural language utterance is part of a conversation that comprises related natural language utterances from the user, wherein determining whether the user has the leader role or the supporter role comprises determining whether the user controls the conversation or whether the user provides information to the speech engine during the conversation.

258. The combination of Smith & Hipp and Guinn discloses a method of *“determining whether the user has the leader role or the supporter role” that “comprises determining whether the user controls the conversation or whether the user provides information to the speech engine during the conversation.”*

259. This claim element essentially has two parts: (1) a requirement for a multi-turn conversation and (2) a requirement for deciding the user’s role based on (a) who (system/user) is directing the dialogue, and/or (b) who is providing information to the other. In other words, the determination of the user’s role (leader vs. supporter) is done by checking whether the system or the user is controlling the flow of the conversation and/or whether the user is providing information in response to the system. Smith & Hipp makes this aspect obvious as well, as it discloses both multi-utterance dialogues and determining role based on dialog control and/or the flow of information between system and user.

260. First, Smith & Hipp is plainly directed to extended voice dialogs rather than “one-shot” or “single-turn” commands – it explicitly discusses complex dialogues composed of multiple related utterances (often organized as “subdialogs”). The reference notes that “efficient human–machine dialog” is “segmented into utterance sequences, *subdialogs*, that are individually aimed at achieving relevant subgoals.” (EX-1005, 282 (emphasis in original).) A given user utterance in Smith & Hipp’s system is thus part of a larger conversation (a sequence of exchanges toward a goal), just as assumed in claim 9. Smith & Hipp also lists among the required dialog “behaviors” “the ability to carry out subdialogs ... and to pass control arbitrarily from one subdialog to another.” (EX-1005, 281.) This indicates the system can handle a conversation that shifts back and forth between the user and the computer in a multi-turn interaction. Even more, Smith & Hipp discloses and describe numerous “sample dialogs” that comprise multi-turn conversations. (EX-1005, 284-85.)

261. Second, Smith & Hipp discloses determining the role of the user based on who is “control[ing] the conversation.” In particular, and as previously discussed, the reference identifies four dialog modes that range from fully system-led (“**directive**”) to fully user-led (“**passive**”), with the defining distinction being who directs the course of the dialog, *i.e.* who controls the conversation. In **directive** mode, “the computer has complete dialog control,” selecting goals and actions while

limiting user input to clarifications. (EX-1005, 282.) By contrast, in **passive** mode, “the user has complete dialog control,” and the system merely “responds directly to user questions” and does not guide the conversation. (EX-1005, 283.) These modes illustrate that the user’s role—as either a leader or a supporter—can be inferred from whether the user is initiating dialog acts and driving the interaction forward (leader), or merely responding to system prompts (supporter).

262. Relatedly, Smith & Hipp explicitly describes a “change of initiative” where the machine should “yield to the user’s control and respond ... to his or her requests.” (EX-1005, 282.) This is precisely determining the user’s role as leader or supporter based on whether the user is controlling the conversation (driving the dialog with requests) or supplying information to the system (following the system’s lead).

263. In addition, Smith & Hipp discloses determining whether the user has a leader or supporter roll based on “*whether the user provides information to the speech engine during the conversation.*” It explains that if the user is more capable or has more relevant knowledge for a given task, the machine will “yield control so that the more competent partner can lead” the dialogue to a solution. (EX-1005, 282.)

264. Smith & Hipp’s **directive** mode vs. **passive** mode illustrates this in practice: in **passive** (user-controlled) mode, the user leads and the system only responds, whereas in **directive** (system-controlled) mode, the system leads and the

user mainly provides info or acknowledges (with any user attempt to control being curtailed except for clarifications). (EX-1005, 282-83.) A POSITA would have found it obvious to implement such a role determination in a voice interface because Smith & Hipp demonstrates that dynamically assigning the leader or supporter role based on who controls the conversation or who is providing whom information greatly improves dialog efficiency and cooperation. (EX-1005, 282.) In fact, Smith & Hipp treats this capability as fundamental to a successful dialog system, identifying at the very beginning of the article that one of the “behaviors required for efficient human-machine dialog” is “the ability to change initiative from strongly computer controlled to strongly user controlled or to some level in between.” (EX-1005, 281.) Therefore, given Smith & Hipp’s clear teaching of multi-turn, mixed-initiative dialogue management, it would have been obvious to a POSITA that determining the user’s role by evaluating if the user is controlling the conversation or providing information (as recited in claim 9) is a straightforward design choice.

265. Guinn is in accord. He describes how initiative (*i.e.* conversation type) depends directly from who is providing information to whom. He writes: “Our model concentrates on two specific dialogue cues: questions and answers. When an agent  $A_1$  asks another agent  $A_2$  to satisfy a goal  $G$ , agent  $A_2$  gains initiative over goal  $G$  and all subgoals of  $G$  until agent  $A_2$  passes control of one of those subgoals back to agent  $A_1$ .” (EX-1006, 279.) A POSITA would understand from this that

determining the leader or supporter roles in a conversation comprises understanding whether the user is providing information to the computer, or vice versa.

266. Thus, the combination of Smith & Hipp and Guinn discloses “[t]he method recited in claim 1, wherein the natural language utterance is part of a conversation that comprises related natural language utterances from the user, wherein determining whether the user has the leader role or the supporter role comprises determining whether the user controls the conversation or whether the user provides information to the speech engine during the conversation.”

## **V. Conclusion**

267. In signing this Declaration, I recognize that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in this proceeding. If required, I will appear for cross-examination at the appropriate time. I reserve the right to offer opinions relevant to the invalidity of the challenged claims at issue and/or offer testimony in support of this Declaration.

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

false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

Dated: April 25, 2025

Respectfully submitted,

A handwritten signature in cursive script that reads "Stuart J. Lipoff". The signature is written in black ink and is positioned above the printed name.

Stuart Lipoff

## APPENDIX

| Exh. | Reference                                                                                                                                                                                    |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001 | U.S. Patent 8,515,765 to Baldwin et al. (“the ’765 patent”)                                                                                                                                  |
| 1002 | Prosecution History for U.S. Patent 8,515,765                                                                                                                                                |
| 1004 | Curriculum Vitae of Stuart Lipoff (“Lipoff CV”)                                                                                                                                              |
| 1005 | Ronnie. W. Smith, et al., <i>An Architecture for Voice Dialog Systems Based on Prolog-Style Theorem Proving</i> , 21(3) Computational Linguistics 281 (1995) (“Smith & Hipp”)                |
| 1006 | Curry I. Guinn, <i>Mechanisms for Mixed-Initiative Human-Computer Collaborative Discourse</i> , Proceedings of the 34th Annual Meeting of the Association for Computational (1996) (“Guinn”) |
| 1007 | Declaration of June Ann Munford in Support of Inter Partes Review of U.S. Patent 8,515,765 (“Munford Declaration”)                                                                           |
| 1008 | Sebastian Möller, <i>Quality of Telephone-Based Spoken Dialogue Systems</i> , Springer (2005) (“Möller”)                                                                                     |
| 1009 | Michael F. McTear, <i>Spoken Dialogue Technology</i> (2004) (“McTear”)                                                                                                                       |
| 1010 | Xuedong Huang et al., <i>Spoken Language Processing</i> , Prentice Hall PTR (2001) (“Huang”)                                                                                                 |
| 1011 | Peter Wiemer-Hastings et al., <i>RMT: A Dialog-Based Research Methods Tutor With or Without a Head</i> , 7th International Conference on Intelligent Tutoring Systems, Vol. 3220 (Jan. 2004) |
| 1012 | Arthur C. Graesser et al., <i>AutoTutor: A Tutor with Dialogue in Natural Language</i> , 36(2) Behavior Research Methods, Instruments, & Computers 180 (2004) (“Graesser”)                   |
| 1013 | Niels Ole Bernsen et al., <i>Designing Interactive Speech Systems: From First Ideas to User Testing</i> , Springer (1998) (“Bernsen”)                                                        |