

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner

v.

DIALECT, LLC,
Patent Owner

Case IPR2024-00753
U.S. Patent No. 7,640,160 B2
Issue Date: December 29, 2009

Title: SYSTEMS AND METHODS FOR RESPONDING TO
NATURAL LANGUAGE SPEECH UTTERANCE

DECLARATION OF CHRISTOPHER M. SCHMANDT

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

TABLE OF CONTENTS

	Page
I. INTRODUCTION AND QUALIFICATIONS	4
A. Qualifications and Experience	4
B. Materials Considered	8
II. PERSON OF ORDINARY SKILL IN THE ART	9
III. STATEMENT OF LEGAL PRINCIPLES	11
A. Claim Construction	11
B. Obviousness (§ 103)	13
IV. THE '160 PATENT	19
A. Overview of the Specification	19
B. The Challenged Claim	21
V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIM	22
A. Brief Summary Of Ground 1 Prior Art	23
1. Kennewick (EX1004)	23
2. Ross (EX1005)	25
B. Ground 1: Claim 12 Is Obvious Over Kennewick In View Of Ross.	32
1. Independent Claim 12: “A method for interpreting natural language utterances using knowledge-enhanced speech recognition engine, wherein the knowledge- enhanced speech recognition engine is configured to determine an intent and correct false recognitions of the natural language utterances, comprising:” (12[pre])	32
(a) “receiving a transcription of a natural language utterance at a computer comprising the knowledge- enhanced speech recognition engine;” (12[a])	34

TABLE OF CONTENTS

(continued)

Page

(b)	“identifying one or more contexts that completely or partially match one or more text combinations contained in the transcription, wherein identifying the matching contexts includes comparing the text combinations against the [sic] grammar expression entries in the [sic] context description grammar and against one or more expected contexts stored in a context stack;” (12[b]).....	44
(c)	“scoring each of the identified matching contexts;” (12[c]).....	63
(d)	“selecting the matching context having a highest score to determine a most likely context for the utterance; and” (12[d]).....	65
(e)	“communicating a request to a domain agent configured to process requests in the most likely context for the utterance, the request formulated using at least one grammar expression entry in the context description grammar.” (12[e])	66
VI.	NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS	68
VII.	CONCLUSION	70

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

I, Christopher M. Schmandt, declare as follows:

I. INTRODUCTION AND QUALIFICATIONS

A. Qualifications and Experience

1. I retired several years ago after a 40-year career at the Massachusetts Institute of Technology (“MIT”); for most of that time I was employed as a Principal Research Scientist at the Media Laboratory. In that role I also served as faculty for the MIT Media Arts and Sciences academic program. I was a founder of the Media Laboratory, a research lab which now spans two buildings.

2. I received my B.S. degree in Electrical Engineering and Computer Science from MIT in 1978, and my M.S. in Visual Studies (Computer Graphics) also from MIT. I was employed at MIT since 1980, initially at the Architecture Machine Group which was an early computer graphics and interactive systems research lab. In 1985, I helped found the Media Laboratory and continued to work there until retirement. I was director of a research group titled “Living Mobile.” My research spanned distributed communication and collaborative systems, with an emphasis on multi-media and user interfaces, with a strong focus on speech-based systems. I have over 70 published conference and journal papers and one book in the field of speech technology and user interaction.

3. For the first fifteen years of my career, my research emphasized speech

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

recognition and speech user interfaces. I built the first conversational computer system utilizing speech recognition and synthesis (“Put That There”) starting in 1980. Put That There parsed spoken input into a semantic structure representing a command or intent, and could ask questions of the user to evoke responses which would be further parsed and melded into the context of the current command. Some information could be gleaned from non-speech channels, specifically where a user was pointing. I continued to innovate speech user interfaces using recognition, text-to-speech synthesis, and recorded audio in a wide variety of projects. I built one of the first graphical user interfaces for audio editing, employing keyword recognition on voice memos in 1982 (“Intelligent Ear”). I built the first research-grade unified messaging system, which combined text and voice messages into a single inbox, with speech recognition over the phone for remote access, and a graphical user interface for desktop access in 1983 (“Phone Slave”). Again, Phone Slave included a parser which could accept fragmentary input and build up a complete representation of the user’s utterance with additional input. Especially in its follow up multi-user system, Conversational Desktop (1985), a number of contextual cues could also aid the dialog understanding, including date, time, calendar entries, and with whom a user was speaking over a teleconference channel. Along with my students we built the first system for real time spoken driving directions, including

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

speech-accessible maps of Cambridge, Massachusetts in 1987 (“Back Seat Driver”). Back Seat Driver had to translate a route from a map form in which each block was a separate street segment with intersection information contained in nodes, into spoken directions, using words like we use with each other to tell the driver where to go. We built some of the earliest speech-based personal assistants for managing messages, calendar, contacts, etc. (Conversational Desktop 1985, Chatter 1993, MailCall 1996). We built quite a few systems employing speech recognition in handheld mobile devices (ComMotion 1999, Nomadic Radio 2000, Impromptu 2001, and Symphony 2004, for example). We applied speech recognition to large bodies of everyday conversations captured with a wearable device and utilized as a memory aid (Memory Prosthesis 2004). We used speech recognition on radio newscasts to build a personalized version of audio newscasts (Synthetic News Radio, 1999) and also investigated adding speech recognition to a mouse-based window system a few years earlier.

4. In addition to this work running my academic research lab, I helped found and then design speech user interfaces for InTouch Systems, which built natural language based personal digital assistants accessible over the telephone. In this context I helped solve a number of problems in obtaining adequate speech recognition performance in the face of personal information, such as contact lists,

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

containing words which were difficult for a system to understand. The system also maintained dialog context, such that while reading email, for example, commands related to email were anticipated, such as finding important messages or messages from some particular person. When a user was accessing a calendar, a different grammar was invoked, emphasizing creation and retrieval of calendar entries by date or time.

5. I was later awarded the prestigious Association for Computing Machinery (ACM) Computer Human Interface (CHI) Academy membership specifically for those years of work pioneering speech user interfaces.

6. In the course of my research, I built a number of speech recognition client/server distributed systems, with the first being in 1985. Much of the initial motivation for a server architecture was that speech recognition required expensive digital signal processing hardware that we could not afford to put on each computer, so a central server with the required hardware was used. Later versions of the speech recognition server architecture allowed certain computers to perform specialized tasks serving a number of client computers providing voice user interfaces, either on screens or over telephone connections. Because of my early work with distributed speech systems, I served for several years in the mid-1990s with a working group on the impact of multimedia systems on the Internet reporting to the Internet

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

Engineering Task Force (IETF) and later the Internet Activities Board (IAB). This work impacted emerging standards such as Session Initiation Protocol (SIP).

7. In my faculty position I taught graduate level courses in speech technology and user interaction design, and directly supervised student research and theses at the Bachelors, Masters, and PhD level. I oversaw the Masters and PhD thesis programs for the entire Media Arts and Sciences academic program during my more senior years. I also served on the Media Laboratory intellectual property committee for many years.

8. My Curriculum Vitae is attached as **Exhibit A**.

9. I have been retained by counsel for Petitioner to provide my expert opinion in connection with the above-captioned proceeding as set forth herein.

B. Materials Considered

10. The analysis that I provide in this Declaration is based on my education, research, and experience, as well as the documents I have considered. In forming my opinions, I have read and considered U.S. Patent No. 7,640,160 B2 (“160 patent”) [EX1001] and its prosecution history. I have cited to the following documents in my analysis below:

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

Exhibit No.	Description of Document
1001	U.S. Patent No. 7,640,160 B2 to Philippe Di Cristo et al. (filed August 15, 2005, issued December 29, 2009) (“ ’160 ” or “ ’160 patent ”)
1004	U.S. Patent App. Pub. No. 2004/0044516 A1 to Robert A. Kennewick et al. (filed June 3, 2003, published March 4, 2004) (“ Kennewick ”)
1005	U.S. Patent App. Pub. No. 2002/0133354 A1 to Steven I. Ross (filed August 16, 2001, published September 19, 2002) (“ Ross ”)
1007	Redline comparison between Kennewick and the ’160 patent
1009	Excerpts from <i>Microsoft Computer Dictionary</i> (5th ed. 2002)
1010	Excerpts from <i>Merriam-Webster’s Collegiate Dictionary</i> (10th ed. 1994)
1012	Prosecution history of U.S. Patent No. 7,640,160 B2

II. PERSON OF ORDINARY SKILL IN THE ART

11. I understand that, under the patent laws in effect before the America Invents Act (“AIA”) of 2011, an assessment of claims of a patent filed before the AIA took effect should be undertaken from the perspective of a person of ordinary skill in the art as of the earliest claimed priority date (i.e., the “time the invention was made”). I have preliminarily for purposes of my Declaration only assumed that date to be August 5, 2005, the filing date of the earliest patent application to which the ’160 patent claims priority (appl. ser. no. 11/197,504).

12. I have also been advised that to determine the appropriate level of a

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

person having ordinary skill in the art, the following factors may be considered: (1) the types of problems encountered by those working in the field and prior art solutions thereto; (2) the sophistication of the technology in question, and the rapidity with which innovations occur in the field; (3) the educational level of active workers in the field; and (4) the educational level of the inventor.

13. The '160 patent states that it “relates to retrieval of information or processing of commands through a speech interface and/or a combination of a speech interface and a non-speech interface.” ('160, 1:8-10.) In my opinion, a person of ordinary skill in the art would have possessed a bachelor's degree in electrical engineering or computer science, and two years of work experience in the field of speech- and natural language-interfaces for interacting with and directing computer systems and software, including speech recognition and processing. A person could also have qualified as a person of ordinary skill in the art with some combination of (1) more formal education (such as a master's of science degree) and less technical experience, or (2) less formal education and more technical or professional experience.

14. My opinions regarding the level of ordinary skill in the art are based on, among other things, my over 40 years of experience in the field of computer science, my understanding of the basic qualifications that would be relevant to an

engineer or scientist tasked with investigating methods and systems in the relevant area, and my familiarity with the backgrounds of colleagues, co-workers, and employees, both past and present.

15. Although my qualifications and experience exceed those of the hypothetical person having ordinary skill in the art defined above, my analysis and opinions regarding the '160 patent have been based on the perspective of a person of ordinary skill in the art as of August 2005.

III. STATEMENT OF LEGAL PRINCIPLES

A. Claim Construction

16. I understand that a purpose of claim construction is to determine what a person of ordinary skill in the art would have understood the claim terms to mean. Claim terms are generally given their ordinary and customary meaning, which is the meaning that the term would have to a person of ordinary skill in the art in question as of the effective filing date.

17. I understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification. I understand that the patent specification, under the legal principles, has been described as the single best guide to the meaning of a claim term, and is thus highly

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

relevant to the interpretation of claim terms. And I understand for terms that do not have a customary meaning within the art, the specification usually supplies the best context of understanding the meaning of those terms.

18. I further understand that other claims of the patent in question, both asserted and unasserted, can be valuable sources of information as to the meaning of a claim term. Because the claim terms are normally used consistently throughout the patent, the usage of a term in one claim can often illuminate the meaning of the same term in other claims. Differences among claims can also be a useful guide in understanding the meaning of particular claim terms.

19. I understand that the prosecution history can further inform the meaning of the claim language by demonstrating how the inventors understood the invention and whether the inventors limited the invention in the course of prosecution, making the claim scope narrower than it otherwise would be. Extrinsic evidence, such as dictionaries, may also be consulted in construing the claim terms.

20. I understand that, in *Inter Partes* Review (IPR) proceedings, a claim of a patent shall be construed using the same claim construction standard that would be used to construe the claim in a civil action filed in a U.S. district court (which I understand is called the “*Phillips*” claim construction standard), including construing the claim in accordance with the ordinary and customary meaning of such

claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

21. I have been instructed by counsel to apply the “*Phillips*” claim construction standard for purposes of interpreting the claims in this proceeding, to the extent they require an explicit construction. The description of the legal principles set forth above thus provides my understanding of the “*Phillips*” standard as provided to me by counsel.

B. Obviousness (§ 103)

22. I understand that a patent claim is obvious if, as of the effective filing date, it would have been obvious to a person having ordinary skill in the field of the technology (the “art”) to which the claimed subject matter belongs.

23. I understand that the following factors should be considered in analyzing obviousness: (1) the scope and content of the prior art; (2) the differences between the prior art and the claims; and (3) the level of ordinary skill in the pertinent art. I also understand that certain other facts known as “secondary considerations” such as commercial success, unexplained results, long felt but unsolved need, industry acclaim, simultaneous invention, copying by others, skepticism by experts in the field, and failure of others may be utilized as indicia of nonobviousness. I understand, however, that secondary considerations should be connected, or have a

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

“nexus,” with the invention claimed in the patent at issue.

24. I understand that a reference qualifies as prior art for obviousness purposes when it is analogous to the claimed invention. The test for determining what art is analogous is: (1) whether the art is from the same field of endeavor, regardless of the problem addressed, and (2) if the reference is not within the field of the inventor’s endeavor, whether the reference still is reasonably pertinent to the particular problem with which the inventor is involved.

25. I understand that a person of ordinary skill in the art is assumed to have knowledge of all prior art. I understand that one skilled in the art can combine various prior art references based on the teachings of those prior art references, the general knowledge present in the art, or common sense. I understand that a motivation to combine references may be implicit in the prior art, and there is no requirement that there be an actual or explicit teaching to combine two references. Thus, one may take into account the inferences and creative steps that a person of ordinary skill in the art would employ to combine the known elements in the prior art in the manner claimed by the patent at issue. I understand that one should avoid “hindsight bias” and ex post reasoning in performing an obviousness analysis. But this does not mean that a person of ordinary skill in the art for purposes of the obviousness inquiry does not have recourse to common sense.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

26. I understand that when determining whether a patent claim is obvious in light of the prior art, neither the particular motivation for the patent nor the stated purpose of the patentee is controlling. The primary inquiry has to do with the objective reach of the claims, and that if those claims extend to something that is obvious, then the entire patent claim is invalid.

27. I understand one way that a patent can be found obvious is if there existed at the time of the invention a known problem for which there was an obvious solution encompassed by the patent's claims. I understand that a motivation to combine various prior art references to solve a particular problem may come from a variety of sources, including market demand or scientific literature. I understand that a need or problem known in the field at the time of the invention can also provide a reason to combine prior art references and render a patent claim invalid for obviousness. I understand that familiar items may have obvious uses beyond their primary purpose, and that a person of ordinary skill in the art will be able to fit the teachings of multiple prior art references together like the pieces of a puzzle. I understand that a person of ordinary skill is also a person of at least ordinary creativity. I understand when there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

technical grasp. If this finite number of predictable solutions leads to the anticipated success, I understand that the invention is likely the product of ordinary skill and common sense, and not of any sort of innovation. I understand that the fact that a combination was obvious to try might also show that it was obvious, and hence invalid, under the patent laws. I understand that if a patent claims a combination of familiar elements according to known methods, the combination is likely to be obvious when it does not more than yield predictable results. Thus, if a person of ordinary skill in the art can implement a predictable variation, an invention is likely obvious. I understand that combining embodiments disclosed near each other in a prior art reference would not ordinarily require a leap of inventiveness.

28. I understand that obviousness may be shown by demonstrating that it would have been obvious to modify what is taught in a single piece of prior art to create the patented invention. Obviousness may also be shown by demonstrating that it would have been obvious to combine the teachings of more than one item of prior art. I understand that a claimed invention may be obvious if some teaching, suggestion, or motivation exists that would have led a person of ordinary skill in the art to combine the invalidating references. I also understand that this suggestion or motivation may come from the knowledge of a person having ordinary skill in the art, or from sources such as explicit statements in the prior art. I understand that

when there is a design need or market pressure, and there are a finite number of predictable solutions, a person of ordinary skill may be motivated to apply common sense and his skill to combine the known options in order to solve the problem.

29. I understand the following are examples of approaches and rationales that may be considered in determining whether a piece of prior art could have been combined with other prior art or with other information within the knowledge of a person having ordinary skill in the art:

- (1) Some teaching, motivation, or suggestion in the prior art that would have led a person of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention;
- (2) Known work in one field of endeavor may prompt variations of it for use in the same field or a different field based on design incentives or other market forces if the variations would have been predictable to a person of ordinary skill in the art;
- (3) Combining prior art elements according to known methods to yield predictable results;
- (4) Applying a known technique to a known device, method, or product ready for improvement to yield predictable results;

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

- (5) Applying a technique or approach that would have been “obvious to try” (choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success);
- (6) Simple substitution of one known element for another to obtain predictable results; or
- (7) Use of a known technique to improve similar products, devices, or methods in the same way.

30. I understand that, when determining whether a claimed combination is obvious, the correct analysis is not whether one of ordinary skill in the art, writing on a blank slate, would have chosen the particular combination of elements described in the claim. Instead, I understand the correct analysis considers whether one of ordinary skill, facing the wide range of needs created by developments in the field of endeavor, would have seen a benefit to selecting the combination claimed.

31. I understand that the test for obviousness is not whether the features of a secondary reference may be bodily incorporated into the structure of the primary reference. The test for obviousness, in other words, is not whether the references could be physically combined but whether the claimed inventions are rendered obvious by the teachings of the prior art as a whole.

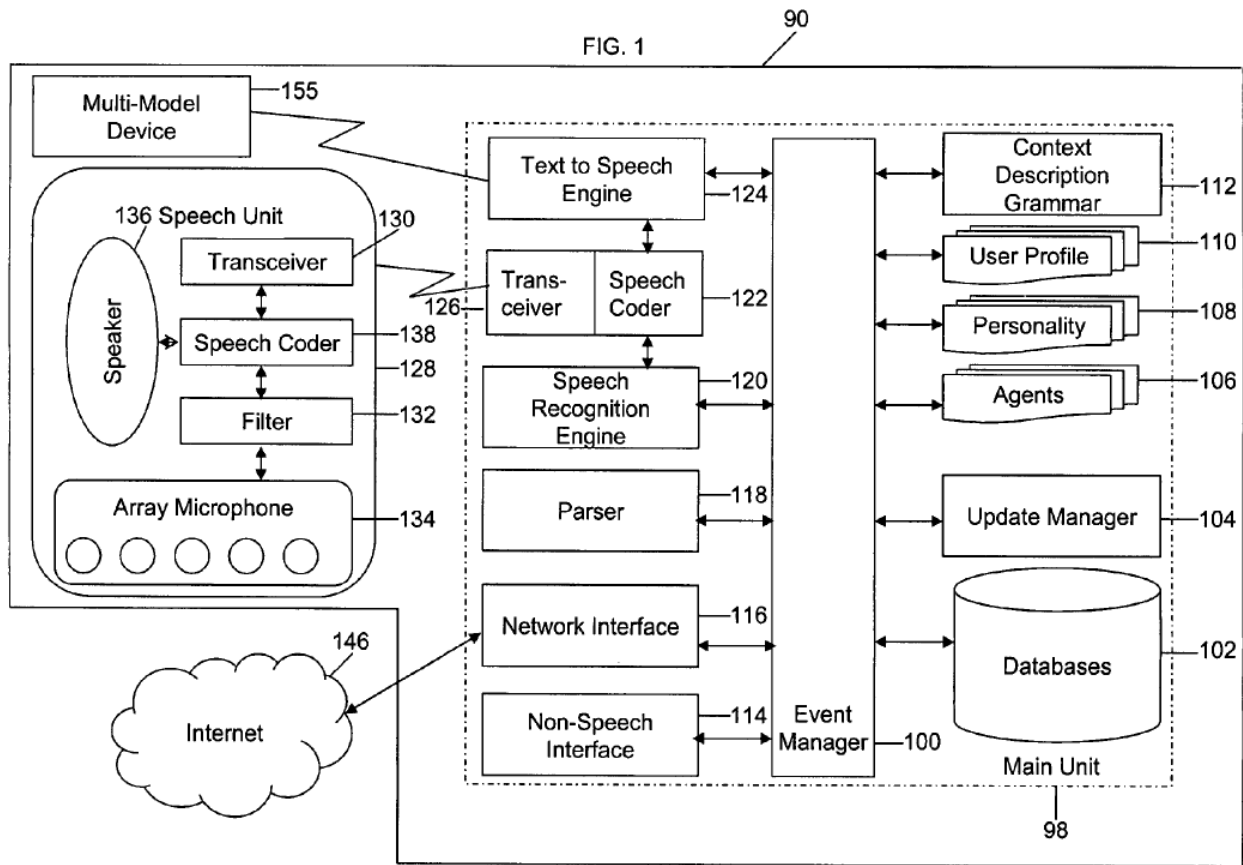
IV. THE '160 PATENT

A. Overview of the Specification

32. The '160 patent states that it “relates to retrieval of information or processing of commands through a speech interface and/or a combination of a speech interface and a non-speech interface.” ('160, 1:8-10.)

33. The '160 patent admits that “[s]peech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications,” and that “[n]atural language processing has been applied to the parsing of speech queries.” ('160, 1:35-38.) According to the '160 patent, however, “current systems do not reliably provide a complete environment for users to submit verbal and/or textual communications through natural language queries that are processed to provide natural responses.” ('160, 1:38-41.) The '160 patent purports to “overcome...drawbacks of prior systems.” ('160, 1:48-50.)

34. Figure 1 of the '160 patent, reproduced below, “is an overall diagrammatic view according to one embodiment of the invention”:



(’160, Fig. 1, 9:63-64.) As shown in the figure above, the system includes, among other things, speech unit **128**, speech recognition engine **120**, context description grammar module **112**, parser **118**, and agents **106**. Speech unit **128** includes a microphone to receive a spoken utterance from a user. (’160, 11:55-57.) After initial processing, the speech is passed to speech recognition engine **120** for processing using the context description grammar module **112**. (’160, 12:23-25.) Then, “[a]ny recognized information may be processed by the parser **118**, which transforms information into complete algorithms and questions using data supplied by

knowledge agents.” (’160, 12:25-28.) “The knowledge agents may then process the commands or questions.” (’160, 12:32-36.) In order to select the correct agent to process the command or question, “[t]he parser **118** uses a scoring system to determine the most likely context or domain for a user’s question and/or command. The score is determined from weighing a number of factors including, the user profile **110**, the domain agent’s data content and previous context. Based on this scoring, the system **90** invokes the correct agent.” (’160, 21:10-15.) The agent returns results to the user. (’160, 12:41-42.)

35. I discuss additional aspects of the ’160 patent as needed in the analysis below.

B. The Challenged Claim

36. This Declaration addresses claim 12 of the ’160 patent, which recites:

12. A method for interpreting natural language utterances using knowledge-enhanced speech recognition engine, wherein the knowledge-enhanced speech recognition engine is configured to determine an intent and correct false recognitions of the natural language utterances, comprising:

- [a] receiving a transcription of a natural language utterance at a computer comprising the knowledge-enhanced speech recognition engine;
- [b] identifying one or more contexts that completely or partially match one

or more text combinations contained in the transcription, wherein identifying the matching contexts includes comparing the text combinations against the grammar expression entries in the context description grammar and against one or more expected contexts stored in a context stack;

- [c] scoring each of the identified matching contexts;
- [d] selecting the matching context having a highest score to determine a most likely context for the utterance; and
- [e] communicating a request to a domain agent configured to process requests in the most likely context for the utterance, the request formulated using at least one grammar expression entry in the context description grammar.

(’160, 30:60-31:14 (Claim 12) (bracketed notation (e.g., “[a]”) added).)

37. I address the challenged claim further in my detailed analysis in **Part V** below.

V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIM

38. I have reviewed and analyzed the prior art references and materials listed in **Part I.B** above. In my opinion, the challenged claim of the ’160 patent would have been obvious to a person of ordinary skill in the art based on the following combination of the prior art.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

Ground	Claims	Basis for Challenge under § 103
1	12	Kennewick (EX1004) in view of Ross (EX1005)

39. I am informed by counsel that each of the references cited in the ground above qualifies as prior art to the challenged claim because each reference was filed and/or published before the '160 patent priority date (August 5, 2005).

A. Brief Summary Of Ground 1 Prior Art

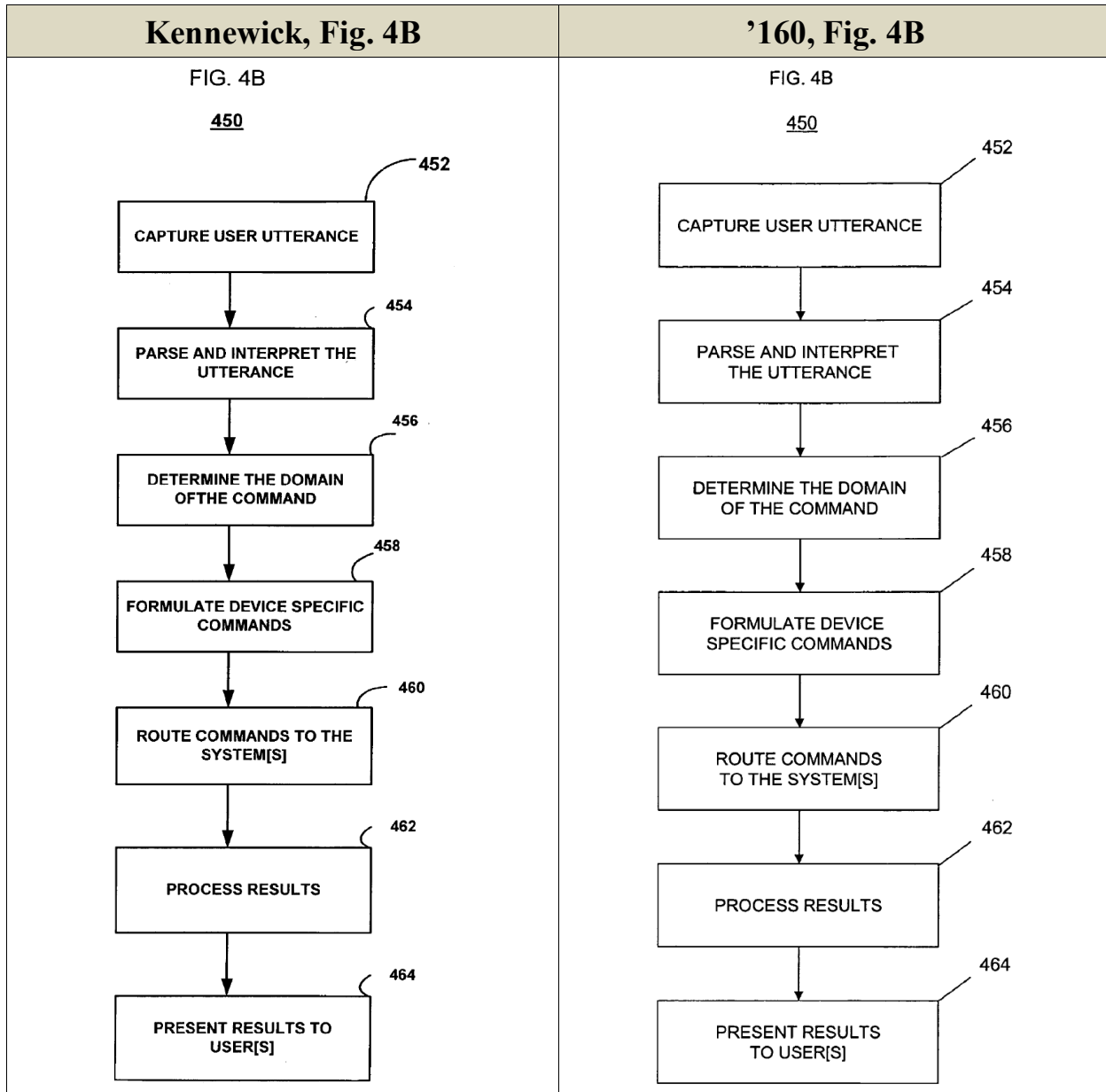
1. Kennewick (EX1004)

40. **Kennewick**, U.S. Patent App. Pub. No. 2004/0044516 A1 to Robert A. Kennewick et al., is entitled “Systems and Methods for Responding to Natural Language Speech Utterance.” I am informed that Kennewick qualifies as prior art because it was filed on June 3, 2003 and published on March 4, 2004, both dates being before the earliest priority date of the '160 patent; in this case, both more than one year before the '160 patent priority date.

41. Kennewick and the '160 patent share a common named inventor, Robert A. Kennewick, and the '160 patent incorporates much of Kennewick nearly verbatim. (*See* EX1007.) As a result, not surprisingly, Kennewick and the '160 patent describe substantially the same subject matter. As shown in my analysis below, Kennewick similarly describes systems and methods for receiving and transcribing a user utterance, processing the text of the transcription using grammars

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

to identify matching contexts, and scoring the contexts in order to select the context and associated application that will be used to process a command/question in the user's utterance. (Kennewick, ¶¶0107-0108, 0142, 0147, 0144-0145, 0152-0161, 0186, 0189.) Figure 4B of Kennewick illustrates an example processing sequence, which as shown below, is identical to Figure 4B of the '160 patent:



42. I discuss additional aspects of Kennewick as needed in my analysis of the challenged claim below.

2. Ross (EX1005)

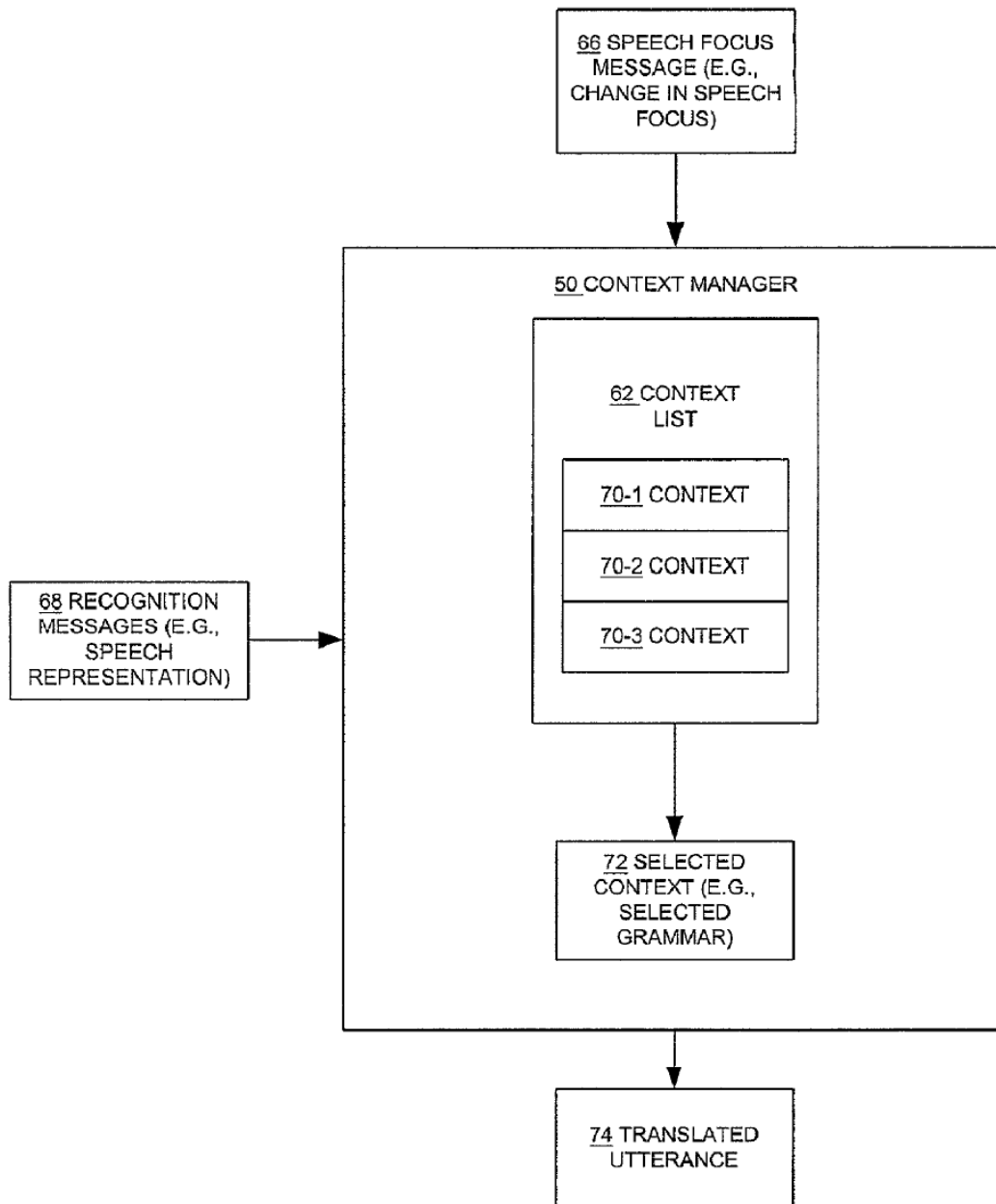
43. Ross, U.S. Patent Application Publication No. 2002/0133354 A1 to

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

Steven I. Ross et al., is entitled “System and Method for Determining Utterance Context in a Multi-Context Speech Application.” I am informed that Ross qualifies as prior art because it was filed on August 16, 2001 and published on September 19, 2002, both dates being before the earliest priority date of the ’160 patent. Ross, as its title suggests, describes techniques for determining a context associated with a user’s spoken command or question in order to determine an application to invoke to process the command or question, just like Kennewick and the ’160 patent. (Ross, e.g., ¶¶0010, 0013.)

44. My analysis points to Ross for its teachings regarding the use of grammars and contexts to select an application to invoke. Claim 12 recites, among other things, a step of “**identifying the matching contexts**” that includes comparing text combinations “**against the [sic] grammar expression entries in the [sic] context description grammar**” as well as “**against one or more expected contexts stored in a context stack.**” As I explain in my analysis below, to the extent that Kennewick alone does not disclose and render obvious these limitations, they would have been obvious in view of Ross.

45. Ross describes a technique where determining a context involves evaluating grammars associated with potential contexts against a user’s recognized speech input—Figure 4 of Ross illustrates a general overview:



(Ross, Fig. 4.) As shown above, Ross includes a context manager **50** that includes a context list **62**. (Ross, ¶0033.) Context list **62**, in turn, “includes contexts **70** (e.g., **70-1**, **70-2**, **70-3**, etc.) for speech-enabled applications **26**, which represent grammars

for the applications 26.¹ (Ross, ¶0033.) A user's transcribed utterance (which Ross calls a "recognition message") is tested against the grammars in the context list in order to determine an application to invoke:

The context manager **50** maintains the priority and state of the various grammars in the context list **62** in the system.... Recognition messages 68 from the speech engine interface 30 are tested by the context manager 50 against the active grammars in the context list 62 in priority order. When a successful match is found, the corresponding translation **74** is dispatched to the script engine **38** for execution, and the priority of the matching grammar (i.e., the selected context **72**) is raised.

(Ross, ¶0034.)

46. Ross further makes clear that "test[ing]...against the active grammars" and finding a successful "match" involves matching text combinations in a user's transcribed utterance to grammar expressions in a grammar. Ross explains that "a grammar is defined for each application **26**" and describes two example grammars, one for an electronic mail application:

¹ All underlining added to quoted material unless noted.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

```
<mail> = do I have any messages |  
    open <message> |  
    create a message |  
    send <message>|  
    print <message>.  
<message> = the? <nth> message | it | this.  
<nth> = first | second | third | fourth | fifth | . . .
```

(Ross, ¶¶0035, 0040) and one for a calendar application:

```
<appointment> = do I have any appointments |  
    open <appointment> |  
    create an appointment |  
    print <appointment>.  
<appointment> = the? <nth> appointment | it | this.
```

(Ross, ¶0046). A grammar expression can generally comprise word tokens, rule references, and grammar operators. For example, in the grammar for the electronic mail application shown above, the line “<message> = the? <nth> message | it | this,” defines a rule named “message.” The rule specifies a grammar expression that includes word tokens (e.g., “the,” “message”), a reference to another rule (“<nth>”) and a grammar operator (“?”). As shown above, each grammar includes a number of grammar expressions. Therefore, as Ross explains, the grammar for the electronic mail application allows spoken phrases such as “open the first message,” “create a message,” “send this,” and “print it” to be matched, while the grammar for the calendar application matches phrases such as “open the first appointment,” “create

an appointment,” “print the fourth appointment,” and “print it.” (Ross, ¶¶0041-0051.) The phrases that can be matched can be easily recognized by substituting the term in angle brackets with text (or expression) to which it can correspond. For example, in the grammar for the electronic mail application, “<message>” can correspond to “the? <nth> message,” “it,” or “this.” Therefore, the grammar expression “print <message>” can match “print it” (as well as “print this” or “print the first message” and so on). Ross discloses, therefore, that a text combination in a user’s transcribed utterance, such as “print it,” matches the “print <message>” expression defined by the “mail” rule as well as the “it” expression defined by the “message” rule in the grammar for the electronic mail application. With respect to the calendar application grammar, “print it” would also match the “print <appointment>” and “it” expressions in the “appointment” rules.

47. Using the example electronic mail and calendar application grammars shown above, Ross further describes how identifying context(s) involves matching transcribed text to the grammar expressions in the grammars:

If both of these grammars were loaded into the context manager **50**, (ignoring foreground/background issues in this example, assuming that the grammars are the same for both foreground and background), the speech center system **20** is listening for any of the phrases accepted by either grammar. If the speech center **20** hears a phrase such as “print

the first message” or “print the first appointment,” the context manager 50 can readily figure out the intended target application 26 for the uttered sentence. Only one grammar will accept the phrase, which thus indicates the selected context 72 for that phrase and that associated application 26 is the one that should be targeted to receive the corresponding command.

If the sentence is “print it” however, both grammars are capable of accepting the utterance. The context manager 50 therefore has to make a choice by referring to the context list 62 of applications in order of recency of access. The context manager 50 tests the utterance against these grammars (indicated by the contexts 70 in the context list 62) in priority order, and passes the commands on to the first application 26 that has a grammar that will accept the phrase.

(Ross, ¶¶0052-0053.)

48. In addition, with respect to context list 62, Ross teaches that the contexts in the list are sorted in priority order, such that the most-recently-accessed context will be the first context evaluated when processing the next user utterance.

(Ross, ¶¶0035 (“[W]henver an application 26 which is not the top priority application 26 is chosen as the target for a speech command, the application 26 indicated by the selected context 72 will move to the head of the list 62. In this way, the last application 26 that the user touched or talked to will get the first opportunity at interpreting the next user utterance, and the other applications 26 will be ordered

in a most-recently-accessed way.”), 0037 (“[T]he context manager **50** receives the recognition message **68** and selects a context **70** for a speech enabled application **26** by checking the speech representation against contexts **70**, starting at top of context list **62** to determine the highest priority context **70** that can process the speech representation”).) Ross’s priority ordering description of a context list is consistent with the well-known concept of a “**stack**,” in which the most recent item added to the stack will be the first item considered from the stack in a later operation. (EX1009, p.305 (definition of “last in, first out”).) This “last in, first out” (“LIFO”) behavior of a stack contrasts with, for example, the “first in, first out” (“FIFO”) behavior of a queue. (EX1009, pp.215 & 305 (definitions of “first in, first out” and “last in, first out”).)

49. I discuss additional aspects of Ross as needed in my analysis of the challenged claim below.

B. Ground 1: Claim 12 Is Obvious Over Kennewick In View Of Ross.

1. Independent Claim 12: “A method for interpreting natural language utterances using knowledge-enhanced speech recognition engine, wherein the knowledge-enhanced speech recognition engine is configured to determine an intent and correct false recognitions of the natural language utterances, comprising:” (12[pre])

50. For reference, claim 12 recites:

12. A method for interpreting natural language utterances using knowledge-enhanced speech recognition engine, wherein the knowledge-enhanced speech recognition engine is configured to determine an intent and correct false recognitions of the natural language utterances, comprising:
- [a] receiving a transcription of a natural language utterance at a computer comprising the knowledge-enhanced speech recognition engine;
 - [b] identifying one or more contexts that completely or partially match one or more text combinations contained in the transcription, wherein identifying the matching contexts includes comparing the text combinations against the grammar expression entries in the context description grammar and against one or more expected contexts stored in a context stack;
 - [c] scoring each of the identified matching contexts;
 - [d] selecting the matching context having a highest score to determine a most likely context for the utterance; and
 - [e] communicating a request to a domain agent configured to process requests in the most likely context for the utterance, the request formulated using at least one grammar expression entry in the context description grammar.

(’160, 30:60-31:14 (Claim 12) (bracketed notation (e.g., “[a]”) added).)

51. The preamble of claim 12 recites “a method for interpreting natural

language utterances using knowledge-enhanced speech recognition engine, wherein the knowledge-enhanced speech recognition engine is configured to determine an intent and correct false recognitions of the natural language utterances, comprising[.]” As I explain in detail for 12[a], Kennewick discloses a “**knowledge-enhanced speech recognition engine**” in the form of a parser that interprets a transcribed natural language utterance by (1) determining the intent of a request in a transcribed natural language utterance to select an application to invoke to process a command or question in the utterance, and (2) identifying potentially false recognitions and prompting the user to correct them. (Kennewick, e.g., ¶¶ 0031, 0092, 0152-0153.) As shown below, the claimed method is disclosed and rendered obvious by the prior art.

(a) “receiving a transcription of a natural language utterance at a computer comprising the knowledge-enhanced speech recognition engine;” (12[a])

52. Kennewick discloses receiving “a **natural language utterance**” from a user. (Kennewick, ¶0010 (“[T]he system comprises a speech unit interface device that receives spoken natural language queries, commands and/or other utterances from a user.”).) With respect to a “**transcription**,” transcribing (i.e. generating a “transcription”) is broadly understood as some form of recording, such as making a copy, whether verbatim or not, or converting information or data from one form to

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

another. (EX1010, p.1253 (entry for “transcribe” stating “**1 a** : to make a written copy of **b** : to make a copy of (dictated or recorded matter) in longhand or on a machine (as a typewriter) **c** : to paraphrase or summarize in writing **d** : WRITE DOWN, RECORD **2 a** : to represent (speech sounds) by means of phonetic symbols **b** : TRANSLATE 2a **c** : to transfer (data) from one recording form to another **d** : to record (as on magnetic tape for later broadcast)...”).) Consistent with this ordinary understanding, the natural language utterance input from the user is passed to a speech recognition engine that generates a “**transcription**” of the utterance by recognizing words and phrases. (Kennewick, ¶¶0142 (“The speech received by the array microphone 134 and passed through the filter **132** may be sent to the speech digitizer or coder 138.”), 0147 (“The output of the speech coder **122** may be fed to the speech recognition engine **120**. The speech recognition engine **120** recognizes words and phrases, using information in the dictionary and phrase tables 112, and passes these to the parser 118 for interpretation.”); *see also id.*, ¶¶0186 (“The process **400** may begin when a user’s utterance (i.e., user query and/or command) is captured through accurate speech recognition operating in a variety of real-world environments at **402**.”), 0187 (“The process 450 may begin when a user’s utterance (i.e., user query and/or command) is captured through accurate speech recognition operating in a variety of real-world environments at 452.”).) The speech recognition

engine's "**transcription**" comprising the recognized words and phrases is then passed to a parser for interpretation. (Kennewick, ¶0147 ("The speech recognition engine 120...passes [recognized words and phrases] to the parser 118 for interpretation."); *see also id.*, ¶¶0152 ("Once the words and phrases have been recognized by the speech recognition engine 120, the tokens and user identification is passed to the parser 118."), ("The decoded speech is processed by the speech recognition engine 120....The recognized words and phrases are processed by the parser 118...").) As explained below, Kennewick's parser corresponds to the claimed "**knowledge-enhanced speech recognition engine**"; and because the parser runs on a computer, the step of receiving the transcription "**at a computer comprising the knowledge-enhanced speech recognition engine**" is disclosed by the parser receiving the transcription.

53. Kennewick's parser corresponds to the claimed "**knowledge-enhanced speech recognition engine**." The written description of the '160 patent describes a "knowledge-enhanced speech recognition engine" in terms of the functionality it provides, and largely tracks the language of claim 12. ('160, 13:62-14:6.) The '160 patent states that the knowledge-enhanced speech recognition engine "may be used to determine the intent of the request and/or to correct false recognitions," that it "may access a set of expected contexts that are stored in a context stack to determine

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

a most likely context” and that it “may use context specific matchers that are able to identify context such as time, location, numbers, dates, categories (e.g., music, movies, television, addresses, etc.) and other context.” (’160, 13:64-14:2.) These descriptions are consistent with the term itself, which suggests software that uses additional/contextual information in connection with performing speech recognition/understanding—for example, to recognize the intent of a user’s request (and thus determine that the speech was likely recognized correctly) or to determine that the recognized speech may be incorrect. I note that the written description of the ’160 patent uses permissive language (e.g., “may”) when describing the knowledge-enhanced speech recognition engine’s functionality and does not require it to be able to perform all, or even any, of the specific functions described. In any event, Kennewick’s parser provides the same functionality. In fact, during prosecution of the ’160 patent, the application identified descriptions of the parser (which also appear in Kennewick) as providing written description support for the “knowledge-enhanced speech recognition engine.” (*Compare* EX1012, p.00281 (“For example, the specification even specifically indicates that ‘the possible contexts are scored and the top one or few are used for further processing’ (pg. 43, paragraph 92). As such, the specification as originally filed does provide support for the features relating to the ‘knowledge-enhanced speech recognition engine’....”))

with Kennewick, ¶0152.)

54. To start, Kennewick's parser is used to determine the intent of a request in the transcribed natural language utterance. It does this by examining the words of the transcribed utterance, as well as by using user profile information and dialog history, to determine a "context." (Kennewick, ¶0152 ("The parser **118** determines a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile **110**, and dialog history.")) "For example," Kennewick explains, "a question with the keywords 'temperature' implies a context value of weather for the question." (Kennewick, ¶0152.) In other words, in this example, the parser determines that the intent of a user's question with the keywords "temperature" is to ask about the weather. By determining the user's intent, the parser can identify the appropriate application and query to execute in order to try to provide the user with the requested information. (Kennewick, ¶¶0152 ("The context of a question or command determines the domain and thereby, the domain agent **156**, if any, to be evoked."), 0092 ("Based on keywords in the questions and commands and the structures of the questions and commands, the parser invokes the required agent[s].")) And by being able to identify the context, the parser also provides an indication that the speech was recognized correctly.

55. Kennewick's parser is also used to correct false recognitions. As I

explain in more detail for 12[c]-[d], Kennewick discloses that the parser “uses a scoring system to determine the mostly likely context or domain for a user’s question and/or command.” (Kennewick, ¶0152.) But “[i]f the confidence level of the score is not high enough to ensure a reliable response, the system 90 may ask the user to verify whether the question and/or command is correctly understood.” (Kennewick, ¶0153.) Kennewick further explains:

If the confidence level of the domain or context score is not high enough to ensure a reliable response, the system can request that the user verify the question or command is correctly understood. In general, the question may be phrased to indicate the context of the question including all criteria or parameters. If the user confirms that the question is correct the system may proceed to produce a response. Otherwise, either the user can rephrase the original question, perhaps adding additional information to remove ambiguity, or the system may ask one or more questions to attempt to resolve the ambiguity or other actions may taken.

(Kennewick, ¶0031.) By causing the system to prompt the user to verify whether the input was correctly understood, the parser provides the ability to correct false recognitions.

56. Next, as I explain in detail for 12[b]-[c], a person of ordinary skill in the art would have understood and found it obvious that Kennewick’s parser would

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

access a set of expected contexts that are stored in a context stack to determine a most likely context. Kennewick discloses the “set of expected contexts” in a few different ways. First, Kennewick states that, as part of determining the most likely context, multiple “possible contexts are scored and the top one or few are used for further processing.” (Kennewick, ¶0152.) Therefore, because these “possible contexts” are recognized as possibly matching the user’s intent, this collection of “possible contexts” is the first way that Kennewick discloses the “set of expected contexts.” Kennewick further discloses that the possible contexts are stored in a context stack, describing an example of how “the contents of the context stack,” such as keywords and criteria corresponding to past contexts, may be used in determining the context of a user’s next request (e.g., “a following-on question”):

[A voice query language] provides a grammar to clearly specify the keyword used to determine the context and a set of one or criteria or parameters....The voice query language may be sensitive to the contents of the context stack. Thus, a following-on question can be asked using an abbreviated grammar, since key words and criteria can be inherited from the stack. For example, the user can simply asked [sic] about another keyword if the criteria of the question remain constant.

(Kennewick, ¶¶0144-0145; *see also id.*, ¶0166 (“Part of the dialog history is maintained in a context stack.”).) In the excerpt above, the context keyword(s)

stored in the context stack and associated with an earlier question is an example of an “expected context.” The set of contexts associated with earlier question(s) is another way that Kennewick discloses and renders obvious the “set of expected contexts.” To the extent there is any question that Kennewick alone discloses and renders obvious accessing expected contexts stored in a context stack in connection with determining a most likely context, it would have been obvious over Ross, as I explain for 12[b].

57. Lastly, Kennewick’s parser also uses “context specific matchers that are able to identify context such as time, location, numbers, dates, categories (e.g., music, movies, television, addresses, etc.) and other context.” (’160, 13:64-14:2.) For example, Kennewick describes “specific parameters or values (criteria) used to determine context for questions and commands.” (Kennewick, ¶0093.) These criteria, like the “context specific matchers” of the ’160 patent, include time, location, dates and other information. (Kennewick, ¶0154 (“For a weather context, examples of criteria include, location, date and time. Other criteria can include command criteria (i.e., yes/no, on/off, pause, stop), and spelling.”); *see also id.*, ¶0189 (describing parameters and criteria).)

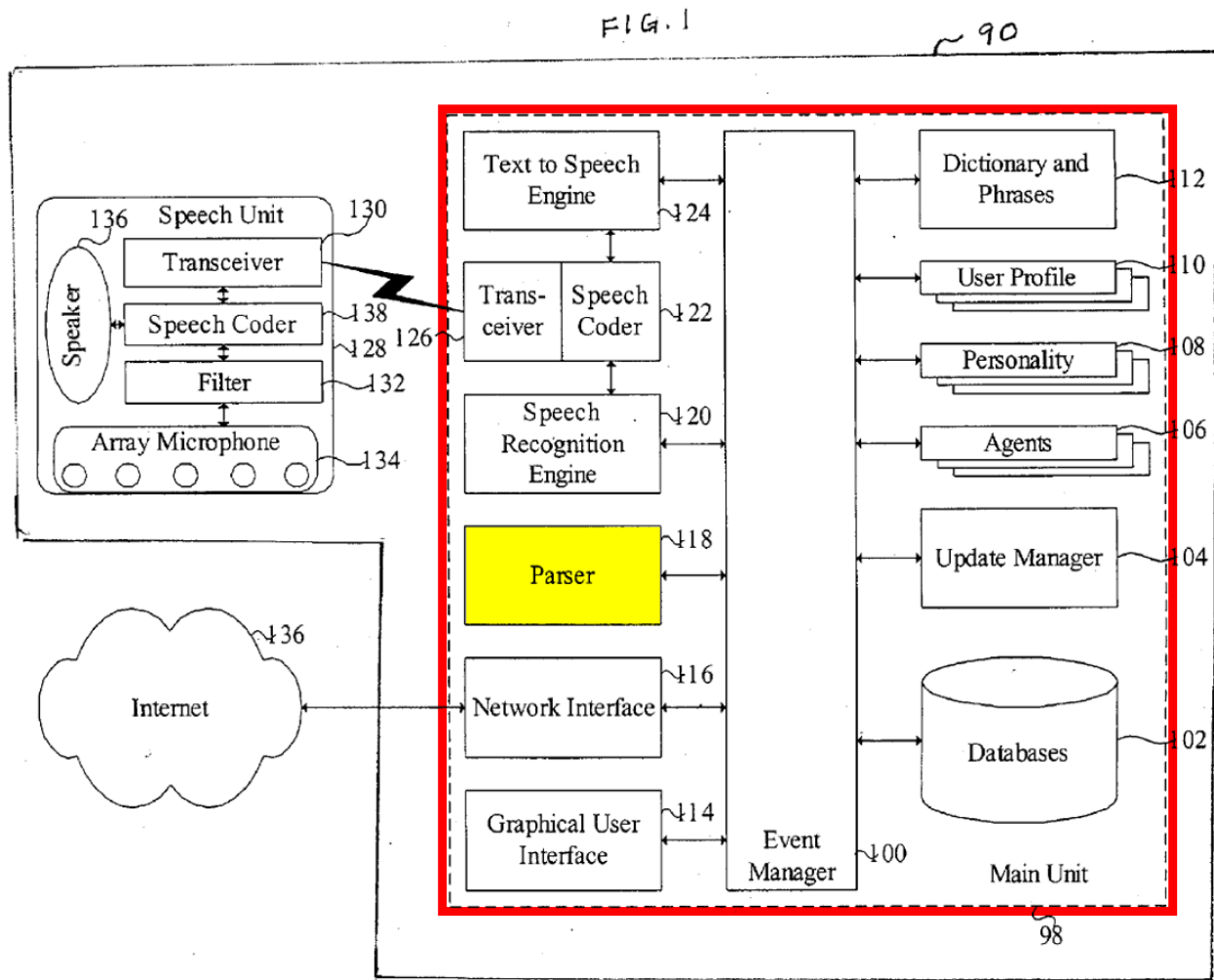
58. Therefore, because Kennewick’s parser includes the specific functionality ascribed to a “**knowledge-enhanced speech recognition engine**” in

the '160 patent's written description—and more generally is software that uses additional/contextual information in connection with performing speech recognition/understanding—Kennewick's parser discloses and renders obvious the claimed **“knowledge-enhanced speech recognition engine.”**

59. Kennewick also discloses a **“computer comprising the knowledge-enhanced speech recognition engine,”** describing how “[o]ne aspect of the invention comprises software that is installed onto the computer, where the software comprises one or more of the following modules:...a parser for parsing the input”:

[T]he computer device or system may comprise a stand alone or networked PC, a PDA or other portable computer device, or other computer device or system. For convenience, these and other computer alternatives shall be simple referred to as a computer. One aspect of the invention comprises software that is installed onto the computer, where the software comprises one or more of the following modules: a speech recognition module for capturing the user input; a parser for parsing the input, a text to speech engine module for converting text to speech; a network interface for enabling the computer to interface with one or more networks; a graphical user interface module and an event manager for managing events.

(Kennewick, ¶0012.) Figure 1 of Kennewick illustrates, depicting a “main unit **98**” (“computer”) that comprises “parser **118**” (“**the knowledge-enhanced speech recognition engine**”):



(Kennewick, Fig. 1 (annotations added).) Therefore, Kennewick's disclosure of the parser receiving a transcription from the speech recognition engine discloses "receiving a transcription of a natural language utterance at a computer comprising the knowledge-enhanced speech recognition engine."

- (b) **“identifying one or more contexts that completely or partially match one or more text combinations contained in the transcription, wherein identifying the matching contexts includes comparing the text combinations against the [sic] grammar expression entries in the [sic] context description grammar and against one or more expected contexts stored in a context stack;” (12[b])**

60. As explained below, 12[b] would have been obvious over Kennewick in view of Ross.

61. As discussed above, Kennewick discloses **“identifying one or more contexts that...match...text in the transcription,”** describing matching tokens (e.g., words) in a recognized utterance to contexts—such as the text “temperature” to the context “weather”:

Once the words and phrases have been recognized by the speech recognition engine 120, the tokens and user identification is passed to the parser 118. The parser 118 examines the tokens for the questions or commands, context and criteria. The parser 118 determines a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, and dialog history. The context of a question or command determines the domain and thereby, the domain agent 156, if any, to be evoked. For example, a question with the keywords “temperature” implies a context value of weather for the question. The parser dynamically receives keyword and associated prior probability or fuzzy possibility updates from the system agent 150

or an already active domain agent **156**. Based on these probabilities or possibilities the possible contexts are scored and the top one or few are used for further processing.

(Kennewick, ¶0152.) I note that, although in this example, the keyword “temperature” does not verbatim match the word “weather” for a weather context, nothing in claim 12 requires this sort of narrow, verbatim matching. Claim 12 merely recites “one or more contexts that completely or partially match one or more text combinations,” which could be satisfied by recognizing that a text combination is associated with a context, and therefore matches the context. In fact, both Kennewick, as shown in the excerpt above, and the ’160 patent state how “[t]he parser **118** determines a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile **110**, and dialog history.” (Kennewick, ¶0152 & ’160, 20:65-67.) A person of ordinary skill in the art would have therefore understood that the text “temperature” **“matches”** the “weather” context. But even if claim 12 required a verbatim match, a person of ordinary skill in the art would have understood and found it obvious, based on Kennewick’s teachings, to identify a context that matches verbatim a text combination in the transcription. For example, it would have been obvious that a keyword “weather” would identify a “weather” context. It would not make sense, for example, for a user’s question, “what’s the weather?” to not match a weather context for processing

the question. Moreover, as I discuss below in connection with the proposed combination based on Ross, Ross discloses verbatim matching of a user's utterance to a context. The combination of Kennewick and Ross therefore also discloses and renders obvious verbatim matching.

62. With respect to “**text combinations**,” in addition to matching a single word, such as keyword “temperature” mentioned in the excerpt above, Kennewick also makes clear that “**text combinations**” are matched, in at least two distinct ways. First, Kennewick discloses examples of not only single-word keywords, but also multi-word keywords, such as “get the value” and “my stock portfolio” in the user utterance, “please get the value of my stock portfolio.” (Kennewick, ¶0186 (“For instance, suppose a user is interested in retrieving the value of her stock portfolio. The user may utter ‘please get the value of my stock portfolio.’ The system 90 may review this request together with stored data such as the user’s profile and determine keywords such as ‘get the value’ and ‘my stock portfolio.’”); *see also id.*, ¶¶0187 (“The user may utter ‘please record my favorite TV program.’ The system 90 may review this request together with stored data such as the user’s profile and determine keywords such as ‘record’ and ‘my favorite TV program.’”), 0098 (“new keywords for a domain, which can include names of politicians, athletes, entertainers, names of new movies or songs, etc. who have achieved recent prominence;”).) A person

of ordinary skill in the art would have understood and found it obvious to identify context(s) by matching multi-word keywords (i.e., “**text combinations**”) in a recognized utterance—for example, to identify a stock portfolio context (and ultimately invoke a stock portfolio agent) by matching the text combinations “get the value” and/or “my stock portfolio.” (Kennewick, e.g., ¶0186 (“a domain [agent] which accesses stock pricing sources to determine the answers to these questions”).)

63. **Second**, Kennewick makes clear that the evaluation of a context can involve matching not only keyword(s) in the transcription, but criteria or parameters in the transcription as well. Kennewick discloses that “[i]n general, the question may be phrased to indicate the context of the question including all criteria or parameters.” (Kennewick, ¶0031.) The criteria or parameters can be “used to determine the context for questions and commands.” (Kennewick, ¶0093 (“The system agent **150** manages the criteria handlers **152**, which handle specific parameters or values (criteria) used to determine context for questions and commands.”).)

64. 12[b] continues by reciting that the claimed “**identifying**” includes: (1) “**comparing the text combinations against the grammar expression entries in the context description grammar**”; and (2) “**comparing the text combinations...against one or more expected contexts stored in a context**

stack.”

65. Turning to the first recited action, to start, I note that 12[b] recites “**the grammar expression entries in the context description grammar**” without any antecedent basis for either “the grammar expression entries” or “the context description grammar.” For purposes of my analysis, I have interpreted “the grammar expression entries in the context description grammar” to mean “**grammar expression entries in a context description grammar.**” As explained below, identifying a context by “**comparing the text combinations against grammar expression entries in a context description grammar**” is disclosed by and obvious over Kennewick, alone or in view of Ross.

66. Kennewick explains that the matching process described above uses a grammar. Kennewick discloses that an agent “passes a regular grammar expression to the parser **118** for evaluation of a context or question.” (Kennewick, ¶0108.) Kennewick further describes how, in one example, a grammar is used “to clearly specify the keyword used to determine the context and a set of one or criteria or parameters.” (Kennewick, ¶0144 (“[T]he voice query language helps users clearly specify the keywords or contexts of the question or command along with the parameters or criteria. The language provides a grammar to clearly specify the

keyword used to determine the context and a set of one or criteria or parameters.”).)²

67. Kennewick also discloses that the grammar would include multiple grammar expressions. Kennewick explains how a “content package” for an agent “include[s] questions or commands” and that each question or command has an associated regular grammar expression:

Content packages include questions or commands. Each command or question or group of commands or questions includes contexts used for creation of one or more queries. The agent **156** passes a regular grammar expression to the parser **118** for evaluation of a context or question. An initial or default context is typically supplied for each command or question. The command or question includes a grammar for the management and evaluation of the context stack.

(Kennewick, ¶0108.)

68. Therefore, a collection of grammar expressions used to identify a context in Kennewick discloses “**grammar expression entries in a context description grammar**” as claimed. The collection discloses “**a context description**

² Kennewick explains that the “voice query language” is intended to assist and improve processing of natural language utterances that the system receives. (Kennewick, ¶¶0144 (“While the invention is intended to be able to accept most any natural language question or command, ambiguity can still be a problem. To assist users formulate concise questions and commands the system **90** may support a voice query language. The language is structured to allow a variety of queries with minimal ambiguity.”), 0032 (similar).)

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

grammar” because it comprises a grammar that describes context(s). This is consistent with the descriptions of a “context description grammar” in the written description of the ’160 patent. For example, the ’160 patent includes a nearly verbatim passage to the one excerpted above from Kennewick, largely simply replacing “regular grammar expression” in Kennewick with “context description grammar expression,” as shown in the side-by-side table below, with underlining added to show differences:

'160, 16:46-54	Kennewick, ¶0108
Content packages include questions or commands. Each command or question or group of commands or questions includes contexts used for creation of one or more requests. The <u>domain</u> agent 156 passes a <u>context description</u> grammar expression to the parser 118 for evaluation of a context or question. An initial or default context is typically supplied for each command or question. The command or question <u>may</u> include a grammar for the management and evaluation of the context stack.	Content packages include questions or commands. Each command or question or group of commands or questions includes contexts used for creation of one or more queries. The agent 156 passes a <u>regular</u> grammar expression to the parser 118 for evaluation of a context or question. An initial or default context is typically supplied for each command or question. The command or question <u>includes</u> a grammar for the management and evaluation of the context stack.

As shown in the excerpt above, the “regular grammar expression” in Kennewick

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

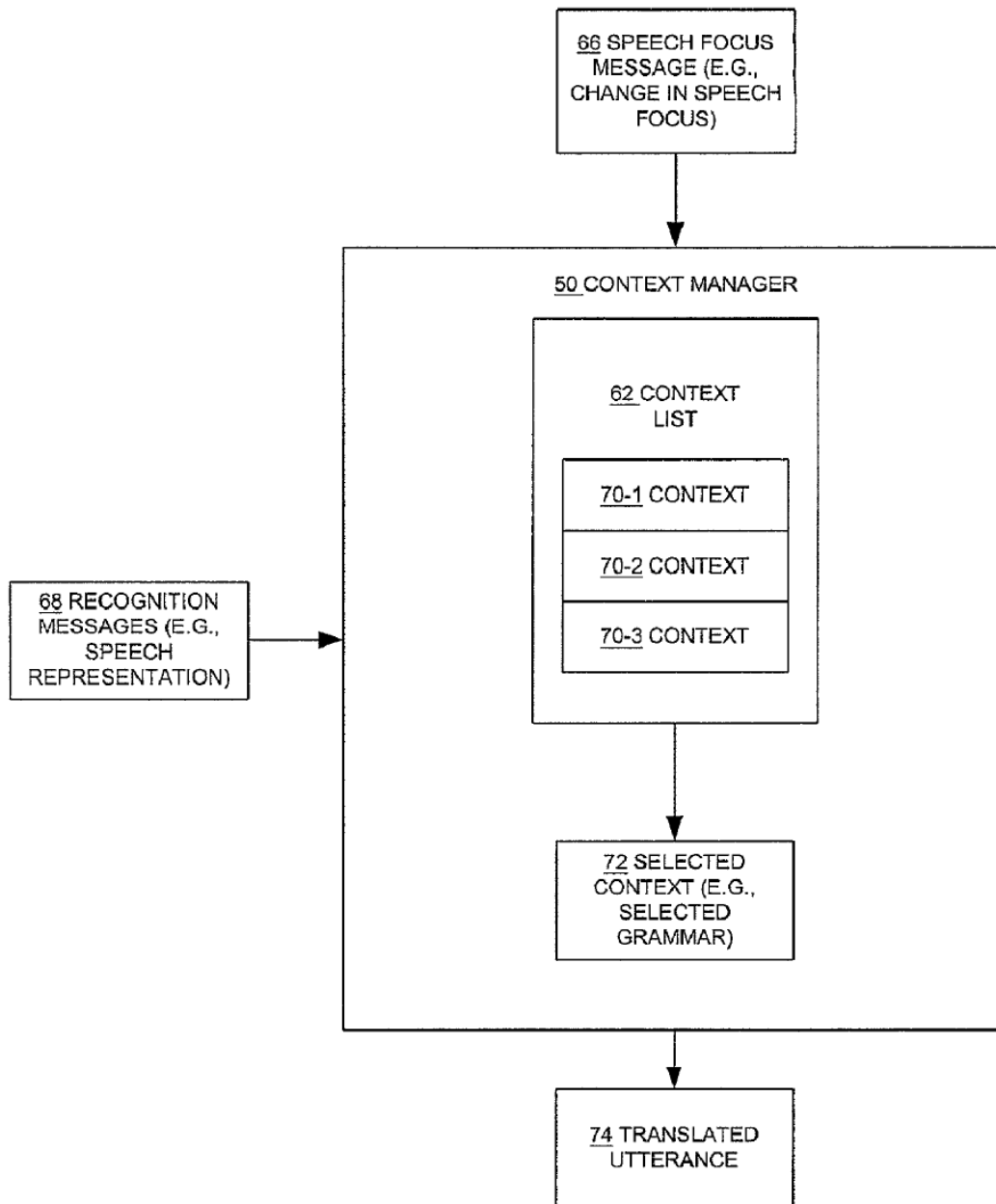
serves the same purpose as the “context description grammar expression” in the ’160 patent: it is passed to the parser for evaluation of a context or question. With respect to evaluating a context with the context description grammar, the written description of the ’160 patent largely tracks the language of claim 12, stating that “[t]he text combination [of a text message] may be compared against entries in a context description grammar that is associated with each agent **106**. If a match is identified between an active grammar in the context description grammar and the command and/or request, then the match may be scored.” (’160, 13:47-50.) Although Kennewick does not also include a nearly-verbatim passage to this one, Kennewick does make clear, and it would have been obvious to a person of ordinary skill in the art, that using a “regular grammar” for evaluation of a context would involve comparing grammar expression entries in the regular grammar to text combinations in a transcribed utterance. This is because, as explained above, Kennewick describes matching text in a transcription, such as keywords and criteria, to grammar expressions to identify context(s) for scoring. (Kennewick, e.g., ¶¶0144 (“[T]he voice query language helps users clearly specify the keywords or contexts of the question or command along with the parameters or criteria. The language provides a grammar to clearly specify the keyword used to determine the context and a set of one or criteria or parameters.”), 0152 (“The parser **118** determines a context for an

utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile **110**, and dialog history.”.) Based on Kennewick’s descriptions, a person of ordinary skill in the art would have found it obvious, for example, that matching “temperature” to identify a “weather” context or matching “stock portfolio” to identify a “stock portfolio” context would involve matching grammar expressions in a grammar.

69. To the extent there is any question that “**grammar expression entries in a context description grammar**” as claimed would have been obvious over Kennewick alone, it would have also been obvious in view of **Ross (EX1005)**.

70. **Ross**, U.S. Patent Application Publication No. 2002/0133354 A1, is entitled “System and Method for Determining Utterance Context in a Multi-Context Speech Application.” (Ross, face page.) Ross, as its title suggests, describes techniques for determining a context associated with a user’s spoken command or question in order to determine an application to invoke to process the command or question, just like Kennewick and the ’160 patent. (Ross, e.g., ¶¶0010, 0013.)

71. Ross describes a technique where determining a context involves evaluating grammars associated with potential contexts against a user’s recognized speech input—Figure 4 of Ross illustrates a general overview:



(Ross, Fig. 4.) As shown above, Ross includes a context manager **50** that includes a context list **62**. (Ross, ¶0033.) Context list **62**, in turn, “includes contexts **70** (e.g., **70-1**, **70-2**, **70-3**, etc.) for speech-enabled applications **26**, which represent grammars for the applications **26**.” (Ross, ¶0033.) A user’s transcribed utterance (a

“recognition message”) is tested against the grammars in the context list in order to determine an application to invoke:

The context manager **50** maintains the priority and state of the various grammars in the context list **62** in the system.... Recognition messages **68** from the speech engine interface **30** are tested by the context manager **50** against the active grammars in the context list **62** in priority order. When a successful match is found, the corresponding translation **74** is dispatched to the script engine **38** for execution, and the priority of the matching grammar (i.e., the selected context **72**) is raised.

(Ross, ¶0034.)

72. Ross further makes clear that “test[ing]...against the active grammars” and finding a successful “match” involves matching text combinations in a user’s transcribed utterance to grammar expressions in a grammar. Ross explains that “a grammar is defined for each application **26**” and describes two example grammars, one for an electronic mail application:

```
<mail> = do I have any messages |  
        open <message> |  
        create a message |  
        send <message>|  
        print <message>.  
<message> = the? <nth> message | it | this.  
<nth> = first | second | third | fourth | fifth | . . .
```

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

(Ross, ¶¶0035, 0040) and one for a calendar application:

```
<appointment> = do I have any appointments |  
    open <appointment> |  
    create an appointment |  
    print <appointment>.  
<appointment> = the? <nth> appointment | it | this.
```

(Ross, ¶0046). A grammar expression can generally comprise word tokens, rule references, and grammar operators. For example, in the grammar for the electronic mail application shown above, the line “<message> = the? <nth> message | it | this,” defines a rule named “message.” The rule specifies a grammar expression that includes word tokens (e.g., “the,” “message”), a reference to another rule (“<nth>”) and a grammar operator (“?”). As shown above, each grammar includes a number of grammar expressions. Therefore, as Ross explains, the grammar for the electronic mail application allows spoken phrases such as “open the first message,” “create a message,” “send this,” and “print it” to be matched, while the grammar for the calendar application matches phrases such as “open the first appointment,” “create an appointment,” “print the fourth appointment,” and “print it.” (Ross, ¶¶0041-0051.) The phrases that can be matched can be easily recognized by substituting the term in angle brackets with text (or expression) to which it can correspond. For example, in the grammar for the electronic mail application, “<message>” can

correspond to “the? <nth> message,” “it,” or “this.” Therefore, the grammar expression “print <message>” can match “print it” (as well as “print this” or “print the first message” and so on). Ross discloses, therefore, that a text combination in a user’s transcribed utterance, such as “print it,” matches the “print <message>” expression defined by the “mail” rule as well as the “it” expression defined by the “message” rule in the grammar for the electronic mail application. With respect to the calendar application grammar, “print it” would also match the “print <appointment>” and “it” expressions in the “appointment” rules.

73. Using the example electronic mail and calendar application grammars shown above, Ross further describes how identifying context(s) involves matching transcribed text to the grammar expressions in the grammars:

If both of these grammars were loaded into the context manager **50**, (ignoring foreground/background issues in this example, assuming that the grammars are the same for both foreground and background), the speech center system **20** is listening for any of the phrases accepted by either grammar. If the speech center **20** hears a phrase such as “print the first message” or “print the first appointment,” the context manager **50** can readily figure out the intended target application **26** for the uttered sentence. Only one grammar will accept the phrase, which thus indicates the selected context **72** for that phrase and that associated application **26** is the one that should be targeted to receive the corresponding command.

If the sentence is “print it” however, both grammars are capable of accepting the utterance. The context manager **50** therefore has to make a choice by referring to the context list **62** of applications in order of recency of access. The context manager **50** tests the utterance against these grammars (indicated by the contexts **70** in the context list **62**) in priority order, and passes the commands on to the first application **26** that has a grammar that will accept the phrase.

(Ross, ¶¶0052-0053.)

74. **Rationale and Motivation to Combine (Kennewick and Ross).** It would have been obvious to a person of ordinary skill in the art to combine Kennewick with Ross’s teachings regarding comparing transcribed user utterances to grammar expressions in a grammar to identify a context. The combination would have predictably resulted in Kennewick’s parser identifying context(s) matching text combination(s) contained in a transcribed user utterance by “**comparing the text combinations against the [sic] grammar expression entries in the [sic; a] context description grammar**” as claimed.

75. Both Kennewick and Ross are analogous references in the same field of speech recognition and processing as the ’160 patent. (’160, 1:8-14; Kennewick, ¶0003; Ross, e.g., ¶¶0010, 0013.) Both Kennewick and Ross would have also been pertinent to a number of problems facing the ’160 patent inventors, including techniques for using a grammar to identify a context for processing a user command

or question and techniques for choosing an application to invoke to process a command or question in a user's transcribed utterance.

76. A person of ordinary skill in the art would have been motivated to combine Kennewick and Ross in the manner described. For starters, as discussed above, a person of ordinary skill in the art would have already understood and found it obvious based on Kennewick's disclosures and background knowledge that Kennewick's use of grammars would involve comparing text combinations against grammar expression entries in a context description grammar as claimed. Ross confirms this understanding by detailing a remarkably similar technique. Moreover, because of the similarity of Ross's technique, a person of ordinary skill in the art, seeking to further understand or implement Kennewick's disclosures, would have appreciated that Ross picks up where Kennewick leaves off and would have turned to Ross for its more detailed teachings.

77. A person of ordinary skill in the art would have had a reasonable expectation of success. The use of grammars and grammar expressions to match text was well-known, and implementing Ross's techniques in Kennewick would have involved the straightforward application of conventional techniques and technologies. For example, Ross explains that the grammars can be specified using the well-known "Backus Naur Form (BNF)." (Ross, ¶0060 ("In a preferred

embodiment, the grammar used for each speech-enabled application **26** may be a Backus Naur Form (BNF) grammar as shown above.”); EX1009, p.49 (dictionary entry for “Backus Naur Form”).) Moreover, both the ’160 patent’s, as well as Kennewick’s, cursory descriptions of grammars confirm that their use was well-known, as I understand that a patent need not teach, and preferably omits, what is well-known in the art. A person of ordinary skill in the art would have found the disclosures of Kennewick and Ross sufficiently detailed to implement the techniques on which I rely, without undue experimentation.

78. 12[b] further recites that identifying the one or more matching contexts also **“includes comparing the text combinations...against one or more expected contexts stored in a context stack.”** As discussed for 12[a], a person of ordinary skill in the art would have understood and found it obvious that Kennewick’s parser would access one or more expected contexts that are stored in a context stack to determine a most likely context. Kennewick discloses the **“one or more expected contexts”** in a few different ways. First, Kennewick states that, as part of determining the most likely context, multiple “possible contexts are scored and the top one or few are used for further processing.” (Kennewick, ¶0152.) Therefore, because these “possible contexts” are recognized as possibly matching the user’s intent, the “possible contexts” are a first way that Kennewick discloses the **“one or**

more expected contexts.” Kennewick further discloses that possible contexts matching a transcription are stored in a context stack, describing an example of how “the contents of the context stack,” such as keywords and criteria corresponding to past contexts, may be used in determining the context of a user’s next request (e.g., “a following-on question”):

[A voice query language] provides a grammar to clearly specify the keyword used to determine the context and a set of one or criteria or parameters....The voice query language may be sensitive to the contents of the context stack. Thus, a following-on question can be asked using an abbreviated grammar, since key words and criteria can be inherited from the stack. For example, the user can simply asked [sic] about another keyword if the criteria of the question remain constant.

(Kennewick, ¶¶0144-0145; *see also id.*, ¶0166 (“Part of the dialog history is maintained in a context stack.”).) In the excerpt above, the context keyword(s) stored in the context stack and associated with an earlier question is an example of an “**expected context.**” The one or more contexts associated with earlier question(s) is another way that Kennewick discloses and renders obvious the “**one or more expected contexts.**” As Kennewick makes clear, the transcribed text of a user’s utterance (e.g., text of “a following-on question”) (“**text combinations**”) can be compared to context keywords in the context stack (“**against one or more expected**

contexts stored in a context stack”).

79. To the extent there is any question that Kennewick alone discloses and renders obvious identifying the one or more matching contexts by “**comparing the text combinations...against one or more expected contexts stored in a context stack,**” it would have been obvious in view of Ross, as noted for 12[a].

80. As discussed above, Ross describes a context manager **50** that includes a context list **62**. (Ross, e.g., Fig. 4, ¶0033.) The context list **62**, in turn, includes contexts **70**. The contexts in the context list **62** are sorted in priority order, such that the most-recently-accessed context will be the first context evaluated when processing the next user utterance. (Ross, ¶¶0035 (“[W]henever an application **26** which is not the top priority application **26** is chosen as the target for a speech command, the application **26** indicated by the selected context **72** will move to the head of the list **62**. In this way, the last application **26** that the user touched or talked to will get the first opportunity at interpreting the next user utterance, and the other applications **26** will be ordered in a most-recently-accessed way.”), 0037 (“[T]he context manager **50** receives the recognition message **68** and selects a context **70** for a speech enabled application **26** by checking the speech representation against contexts **70**, starting at top of context list **62** to determine the highest priority context **70** that can process the speech representation”).) Ross’s priority ordering description

of a context list is consistent with the well-known concept of a “stack,” in which the most recent item added to the stack will be the first item considered from the stack in a later operation. (EX1009, p.305 (definition of “last in, first out”).) This “last in, first out” (“LIFO”) behavior of a stack contrasts with, for example, the “first in, first out” (“FIFO”) behavior of a queue. (EX1009, pp.215 & 305 (definitions of “first in, first out” and “last in, first out”).)

81. **Rationale and Motivation to Combine (Kennewick and Ross).** I discussed the rationale and motivation to combine Kennewick and Ross, which applies here as well. Moreover, with respect to this portion of 12[b], it would have been obvious to combine Kennewick with Ross’s disclosures regarding storing contexts associated with accessed applications on a context stack. The combination would have predictably resulted in “**one or more expected contexts**”—such as the contexts associated with applications that had already been invoked—“**stored in a context stack.**”

82. A person of ordinary skill in the art would have been motivated to combine. The combination would have involved the application of known techniques to improve a similar system in the same way. As discussed above, Kennewick already describes using a context stack storing context keywords in connection with determining a context for a user’s question. Kennewick further

suggests that the context keywords in the context stack correspond to previously identified contexts, by describing how “key words and criteria can be inherited from the stack” when evaluating “a following-on question.” (Kennewick, ¶¶0144-0145.) Ross, with a similar system, confirms storage of “**expected contexts**” in a “**context stack**,” and that doing so enables comparing text against the expected contexts to efficiently identify a matching context. (Ross, ¶0035 (“The speech focus priority is established by maintaining an ordered list of applications (e.g., context list **62**). Whenever an application **26** gains window focus, it will move to the head of the list **62**. Likewise, whenever an application **26** which is not the top priority application **26** is chosen as the target for a speech command, the application **26** indicated by the selected context **72** will move to the head of the list **62**. In this way, the last application **26** that the user touched or talked to will get the first opportunity at interpreting the next user utterance, and the other applications **26** will be ordered in a most-recently-accessed way.”).)

(c) “**scoring each of the identified matching contexts;**”
(12[c])

83. Kennewick discloses “**scoring each of the identified matching contexts.**” As discussed for 12[b], Kennewick describes identifying possible contexts that match a user’s utterance “by applying prior probabilities or fuzzy possibilities to keyword matching, user profile **110**, and dialog history.”

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

(Kennewick, ¶0152.) “Based on these probabilities or possibilities the possible contexts are scored and the top one or few are used for further processing.”

(Kennewick, ¶0152; *see also id.*, ¶¶0153 (“The parser **118** uses a scoring system to determine the mostly likely context or domain for a user’s question and/or command.”), 0030 (“The system may determine the mostly likely context or domain for a user’s question or command, for example, by using a real-time scoring system or other technique.”).)

84. I further note that, during prosecution of the ’160 patent, the applicant identified a substantially verbatim passage in the ’160 application as supporting the written description disclosure of this limitation. (EX1012, pp.00275 (pending claim 45 reciting “scoring each of the identified matching contexts”), 00280-00281 (applicant response purporting to identify written description support).) The applicant stated:

The specification further describes how when a matching context is identified, “then the match may be scored,” wherein domain agents associated with various respective contexts may be selected or otherwise “ranked based on the determined score.” The scoring process used at various stages of a dialog is described throughout the specification (e.g., pgs. 10-11, paragraphs 23-24, pg. 42, paragraph 91, pg. 46, paragraph 101, etc.). For example, the specification even specifically indicates that “the possible contexts are scored and the top

one or few are used for further processing” (pg. 43, paragraph 92). As such, the specification as originally filed does provide support for the features relating to the “knowledge-enhanced speech recognition engine,” including identifying matching contexts, scoring the matching contexts, and selecting a context with the highest score as the “most likely” context.

(EX1012, p.00281.)

(d) **“selecting the matching context having a highest score to determine a most likely context for the utterance; and” (12[d])**

85. As discussed for 12[c], Kennewick discloses that “the possible contexts are scored and the top one or few are used for further processing.” (Kennewick, ¶0152.) Kennewick further describes how “[t]he parser **118** uses a scoring system to determine the mostly likely context or domain for a user’s question and/or command” and that “[b]ased on this scoring, the system **90** invokes the correct agent.” (Kennewick, ¶0153.) A person of ordinary skill in the art would have therefore understood and found it obvious that the matching context **“having a highest score”** would determine the **“most likely context”**—first, because Kennewick discloses selecting only the context with the “top” score, and second, because it would be quite sensible to select the context having a highest score as a mostly likely context even where “the top...few are used for further processing.”

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

86. Moreover, as with 12[c], during prosecution of the '160 patent, the applicant identified a passage in the '160 application that is substantially verbatim paragraph 0152 in Kennewick as supporting the written description disclosure of this limitation. (EX1012, pp.00275 (pending claim 45 reciting “selecting the matching context having a highest score to determine a most likely context for the utterance”), 00280-00281 (applicant response purporting to identify written description support, include paragraph 92 reciting “the possible contexts are scored and the top one or few are used for further processing”).)

(e) **“communicating a request to a domain agent configured to process requests in the most likely context for the utterance, the request formulated using at least one grammar expression entry in the context description grammar.” (12[e])**

87. Kennewick discloses **“communicating a request to a domain agent configured to process requests in the most likely context for the utterance.”** Kennewick explains, for example, that “[b]oth system agents **150** and domain agents **156** receive questions and commands from the parser **118**” and that “[a]fter selecting the agents **106**, a properly formatted query and/or command that will be submitted to the agents **106** may be determined at **608**.” (Kennewick, ¶¶0092, 0189; *see also id.*, ¶¶0091 (describing domain agents), 0185 (“Once, the properly formatted queries and/or commands are generated, the appropriate action may be taken at **308**. This

may require that the properly formatted queries and/or commands be sent to specific domain agent[s]...that can fulfill the requirements of the query and/or command.".)

Kennewick therefore teaches "**communicating a request**" (submitting a "properly formatted query and/or command" from the parser) "**to a domain agent configured to process requests in the most likely context for the utterance**" (selected domain agent(s) **156**).

88. Kennewick further discloses and renders obvious that the request communicated to a domain agent would be "**formulated using at least one grammar expression entry in the context description grammar.**" As discussed above for 12[b], Kennewick discloses and renders obvious a context description grammar that includes one or more grammar expression entries. As noted, Kennewick refers to a "regular grammar" that serves the same purpose of the "context description grammar" in the '160 patent. (*E.g.*, Kennewick, ¶0108; '160, 16:46-54.) As it relates to 12[e], Kennewick refers back to the "regular grammar," similarly disclosing "formulat[ing] a question or command in the regular grammar used by agents." (Kennewick, ¶0154 ("To formulate a question or command in the regular grammar used by agents, the parser **118** may determine required and optional values for the criteria or parameters."); *see also id.*, ¶0189 ("In order to formulate a question or command in the regular grammar used by agents **106**, a determination

may be made as to the required and optional values for the criteria or parameters of the queries and/or commands.”.) Kennewick, alone or in view of Ross, therefore discloses formulating a request to a domain agent **“using at least one grammar expression entry in the context description grammar.”** Moreover, because grammar expression entries are used when matching a transcription to determine a context and application to invoke, as taught by both Kennewick and Ross, a person of ordinary skill in the art would have understood that the resulting request communicated to the domain agent would have been formulated **“using at least one grammar expression entry in the context description grammar.”**

VI. NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

89. As explained in **Part III.B**, I understand that “secondary considerations” may be used as indicia of nonobviousness, provided that any such consideration has a nexus to the invention claimed. I understand from Petitioner’s counsel that Patent Owner in the related district court litigation has not yet identified any evidence related to secondary considerations with a nexus to the claimed invention. Nor am I aware of information, such as commercial success, unexplained results, long felt but unsolved need, industry acclaim, simultaneous invention, copying by others, skepticism by experts in the field, and failure of others, suggesting that the claim addressed in this Declaration are not obvious. To the extent

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,640,160 B2

Patent Owner later provides information it claims relates to secondary considerations, I reserve the right to supplement my analysis and opinions to comment on it.

VII. CONCLUSION

90. 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 claim and/or offer testimony in support of this Declaration.

91. 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 4, 2024

Respectfully submitted,

A handwritten signature in black ink, appearing to read 'Chris Schmandt', written over a horizontal line.

Christopher M. Schmandt
Winchester, MA

Exhibit A

Christopher Schmandt

4 Longfellow
Winchester, MA 01890
+1-617-230-4257

Education

M.I.T., Master of Science, Visual Studies (Computer Graphics), 1980

M.I.T., Bachelor of Science, Computer Science, 1978

Professional Experience -- MIT

Media Laboratory, Principal Research Scientist, 1985-2018 (retired)
Director, Living Mobile Research Group (formerly Speech + Mobility)

Architecture Machine Group, Research Associate, 1980-1984

Architecture Machine Group, Research Assistant, 1979-1980

Architecture Machine Group, Graphics Programmer, 1977-1979

Departmental Undergraduate Research Opportunities Program Coordinator, 1984-2018

Laboratory Intellectual Property Committee 2001-2016, chair 2002-2009

Departmental Committee on Graduate Studies, 1996-2001, 2007-2018

Sponsored Research Activities

Alerting and Mobile Messaging, Digital Life Consortium, MIT Media Lab, 1997-2018

Acoustical Cues to Discourse Structure, National Science Foundation, Principal Investigator, 1995-1998

Parsing Radio, News in the Future Consortium, MIT Media Lab, 1993-1999

Desktop Audio, SUN Microsystems, Inc., Principal Investigator, 1989-1996

Voice Interaction in Hand Held Computers, Apple Computer, Principal Investigator, 1991-1993

Voice Interfaces for Network Services, AT&T, Principal Investigator, 1989-1991

Back Seat Driver, NEC, Principal Investigator, 1988-1991

Acoustic and Visual Cues for Speech Recognition, DARPA, co-Principal Investigator, 1986-1988

Personal Computers and Telephony, NTT Public Corporation, Principal Investigator, 1984-1989

Home Telecomputing, Atari, Inc., Principal Investigator, 1983

Testifying legal engagements within the last five years, representing party in italics

- *Microsoft v. SpeakWare*, 2019
- *Bumble Trading LLC v. Match Group LLC*, 2019
- *Tile Inc. v. Cellwitch Inc.*, 2019
- *Blackberry Ltd v. Facebook, Inc* 2020
- *EROAD Ltd v. PerDiem Co.*, 2020
- *Snap Inc. v. SRK Technology LLC*, 2020
- *Shopify Inc Ltd v. Express Mobile Inc.*, 2020
- *Express Mobile Inc. v Wix Ltd* 2020
- *Motorola Mobility LLC v. Ironworks patents, LLC.* 2021
- *Express Mobile Inc. v GoDaddy Inc* 2021
- *Bumble Trading LLC v. Kinectus, LLC.* 2021
- *Tile, Inc v. Linquet Technologies, Inc.* 2021
- *Facebook, Inc. v. Palo Alto Research Center, Inc.* 2021
- *Unified Patents, LLC v. Gesture Technology Partners, LLC.* 2021
- *Quantum Metric, Inc v. Content Square Isreal, Ltd.* 2021
- *Meta Platforms Inc. v. AlmondNet, Inc* 2022
- *Apple Inc. v. Zentian Limited* 2022
- *Google LLC v. Mira Advanced Technology Systems Inc* 2022
- *Google LLC v. Robocast* 2023
- *Google LLC v. Buffalo Patents LLC* 2023
- *Amazon v. AlmondNet* 2023
- *Google LLC v. Valtrus* 2023
- *Samsung Electronics Co., Ltd v. Stanton Techiya, LLC* 2024
- *Roku, Inc. v. AlmondNet, Inc.* 2024
- *LinkedIn v. AlmondNet, Inc* 2024

Professional Experience -- Intellectual Property

Co-inventor on U.S. Patents

- 5,177,685 for "Automobile navigation system using real time spoken driving instructions"
- 6,728,348 for "System for storing voice recognizable identifiers using a limited input device such as a telephone key pad"
- 6,937,986 for "Automatic dynamic speech recognition vocabulary based on external sources of information"
- 7,098,776 for "Methods and apparatus for vibrotactile communication"
- 7,392,280 for "Method for summarization of threads in electronic mail"
- 7,738,637 "Interactive voice message retrieval"
- 7,443,283 for "Methods and apparatus for connecting an intimate group by exchanging awareness cues and text, voice instant messages, and two-way voice communications"
- 7,865,560 for "System for summarization of threads in electronic mail"
- 8,121,653 for "Methods and apparatus for autonomously managing communications using an intelligent intermediary"
- 8,135,128 for "Animatronic creatures that act as intermediaries between human users and a telephone system"

Co-inventor on U.S. Patent Applications

- 20020076009 "International dialing using spoken commands"
- 20020087328 "Automatic dynamic speech recognition vocabulary based on external sources of information"

- 20020064257 "System for storing voice recognizable identifiers using a limited input device such as a telephone keypad"
- 20030023688 "Voice-based message sorting and retrieval method"
- 20030081738 "Method and apparatus for improving access to numerical information in voice messages"
- 20030144846 "Method and system for modifying the behavior of an application based upon the application's grammar"
- 20030158903 "Calendar bar interface for electronic mail interaction"

Publications

Field Study of a Tactile Sound Awareness Device for Deaf and Hard of Hearing Users ISWC 2020. (with D. Jain, B. Chiu, S. Goodman, L. Findlater, J. Froehlich)

SkinMorph: Texture-Tunable On-Skin Interface Through Thin, Programmable Gel ISWC 2018. (with Cindy Kao, M. Banforth, D. Kim)

Technical Interventions to Detect, Communicate, and Deter Sexual Assault. ISWC 2017. (with Manisha Mohan)

Exploring Interactions and Perceptions of Kinetic Wearables. DIS 2017. (with Cindy Hsin-Liu Kao, D. Ajilo, O. Anilonyte, A. Dementyev, I. Choi and S. Follmer)

Leveraging User-made Predictions to Help Understand Personal Behavior. MobileHCI 2017. (with Miriam Greis, Tilman Dangler and Albrecht Schmidt)

Rovables: On-Body Robots as Mobile Wearables. UIST 2016. (with Cindy Hsin-Liu Kao, A. Dementyev, I. Choi, D. Ajilo, M. Xu, and S. Follmer)

DuoSkin: Rapidly Prototyping On-Skin User Interfaces Using Skin-Friendly Materials. ISWC 2016. (with Cindy Hsin-Liu Kao, Christian Holz, Asta Roseway, and Andres Calvo)

Immersive Terrestrial Scuba Diving Using Virtual Reality (with Dhruv Jain, Misha Sra, Jingru Go, Rodrigo Margues, Raymond Wu and Justin Chiu) Proceedings, UIST 2016

Expanding social mobile games beyond the device screen (with Misha Sra) Journal of Personal and Ubiquitous Computing, 2015

NailO: Fingernails as an input surface (with Cindy Hsin-Liu Kao, Artem Dementyev, Joseph Paradiso) CHI 2015

Mugshots: A mug display for front and back stage social interaction in the workplace (with Cindy Hsin-Liu Kao) TEI (Tangible and Embedded Interfaces) 2015

Mime: compact, low power 3D gesture sensing for interaction with head mounted displays (with Andrea Colaco, Ahmend Kirmani, Hye Soo Yang, Nan-Wei Gong, and Vivek Goyal) Proceedings of UIST 2013.

Spotz: A location-based approach to self-awareness (with Misha Sra) Proceedings of Persuasive 2013.

Setting the stage for interaction: A tablet application to augment group discussion in a seminar class (with Drew Harry and Eric Gordon) Proceedings of CSCW 2012.

Indoor Location Sensing using Geo-Magnetism (with Jaewoo Chung, Matt Donahoe, Ig-Jae Kim, Pedram Razavai and Micaela Wiseman) Proceedings of International Conference on Mobile Systems, Applications, and Services (Mobisys) 2011.

My second bike: a TV-enabled social and interactive riding experience (with Jaewoo Chung, Kuang Xu, Andrea Colaco, and Victor Li) Proceedings of IEEE Communications and Networking Conference, Jan 2010.

Going my way?: User-aware route planner (with Jaewoo Chung) Proceedings CHI 2009.

Globetoddler: Designing for remote interaction between preschoolers and their traveling parents (with Paulina Modlitba) CHI 2008 Extended Abstracts

Are we there yet? - a temporally aware media player (with Matt Adcock and Jaewoo Chung), Australian User Interface Conference (AUIC) 2008

Physical embodiments for mobile communication agents (with Stefan Marti), UIST 2005

Giving the caller the finger: collaborative responsibility for cellphone interruptions (with Stefan Marti) Extended Abstracts, CHI 2005

Active Messenger: email filtering and delivery in a heterogeneous network (with Stefan Marti) Human-Computer Interaction Journal (HCI) Volume 20 (2005)

WatchMe: communication and awareness between members of a closely-knit group (with Natalia Marmasse) Proceedings of Ubicomp 2004

An audio-based personal memory aid (with S. Vemuri, W. Bender, S. Tellex and B. Lassey) Proceedings of Ubicomp 2004

Improving speech playback using time-compression and speech recognition (with Sunil Vemuri, Philip DeCamp, and Walter Bender) Proceedings of CHI 2004

Impromptu: managing networked audio applications for mobile users (with Kwan Lee, Jang Kim, and Mark Ackerman), Proceedings of MobiSys 2004

TalkBack: a conversational answering machine (with Vidya Lakshminpathy and Natalia Marmasse) Proceedings of UIST 2003

"ListenIn" to domestic environments from remote locations (with Gerardo Vallejo) Proceedings of the 2003 International Conference on Auditory Display (ICAD)

Safe & Sound: a wireless leash (with Natalia Marmasse) Extended Abstracts, Proceedings of CHI 2003

Mediated voice communication via mobile IP (with Jang Kim, Kwan Lee, Gerardo Vallejo, and Mark Ackerman), Proceedings of UIST 2002.

The Audio Notebook: Paper and pen interaction with structured speech (with Lisa Stifelman and Barry Arons), Proceedings of CHI 2001

Synthetic News Radio (with Keith Emnett) IBM Systems Journal, Vol. 39 Nos. 3-4, pp. 646-659, 2000.

Everywhere messaging (with Natalia Marmasse, Stefan Marti, Nitin Sawhney, and Sean Wheeler) IBM Systems Journal, Vol. 39 Nos. 3-4, pp. 660-677, 2000.

Location-aware information delivery with comMotion (with Natalia Marmasse), Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing, pp. 157-171, Springer, 2000.

Nomadic Radio: Scalable and contextual notification for wearable audio messaging (with Nitin Sawhney), Proceedings of CHI 1999.

Speaking and listening on the run: Design for wearable audio computing (with Nitin Sawhney), Proceedings of International Symposium on Wearable Computing, 1998.

Audio Hallway: A virtual acoustic environment for browsing, Proceedings of UIST 1998.

Dynamic Soundscape: Mapping time to space for audio browsing (with Minoru Kobayashi), Proceedings of CHI 1997.

CLUES: Dynamic personalized message filtering (with Matt Marx), Proceedings of CSCW 1996.

Using acoustic structure in a hand-held audio playback device (with Deb Roy), IBM Systems Journal, Vol 35, Nos. 3 and 4, 1996.

Mailcall: Message presentation and navigation in a nonvisual environment (with Matt Marx), Proceedings of CHI 1996

AudioStreamer: Exploiting simultaneity for listening (with Atty Mullins), short paper, CHI 1995.

Multimedia momadic services on today's hardware, IEEE Network, September/October 1994.

Putting people first: Specifying proper names in speech interfaces (with Matt Marx), proceedings of UIST 1994.

Chatter: A conversational learning speech interface (with E. Ly) AAAI Workshop on Intelligent Multi-Media Multi-Modal Systems, 1994.

Voice Communication with Computers: Conversational Systems. New York: Van Nostrand Reinhold. 1994.

Capturing, structuring, and representing ubiquitous audio (with D. Hindus and C. Horner), ACM Transactions on Information Systems, Vol. 11, No. 4, October 1993.

Speech Recognition Architectures for Multimedia Environments, (with E. Ly and B. Arons), Proceedings of the 1993 AVIOS Conference, September 1993.

Phoneshell: the Telephone as Computer Terminal, Proceedings of the ACM Multimedia Conference, August 1993.

Voicenotes: A Speech Interface for a Hand-Held Voice Notetaker (with L. Stifelman, B. Arons, and E. Hulteen), Proceedings of INTERCHI'93, April 1993.

From Desktop Audio to Mobile Access: Opportunities for Voice in Computing, book chapter in Advances in Human-Computer Interaction, Vol. 4, H.R. Hartson and D. Hix editors. 1992.

Ubiquitous Audio: Capturing Spontaneous Collaboration (with D. Hindus), Proceedings of CSCW'92, November 1992.

Integrating Audio and Telephony in a Distributed Workstation Environment (with S. Angebrannt, R. Hyde, D. Luong, and N. Siravara), Proceedings of the Summer 1991 USENIX Conference, June 1991.

Augmenting a Window System with Speech Input (with M. Ackerman and D. Hindus), Computer, IEEE Computer Society, Vol. 23, No. 8, August 1990.

Observations on Using Speech Input for Window Navigation (with D. Hindus, M. Ackerman, and S. Manandhar), Proceedings, Human-Computer Interaction, Interact '90, IFIP, August 1990.

Phonetool: Integrating Telephones and Workstations (with S. Casner), Proceedings, GLOBECOM '89, IEEE Communications Society, November 1989.

Desktop Audio (with B. Arons), UNIX Review, October 1989.

Synthetic Speech for Real Time Direction-Giving (with J. Davis), IEEE Transactions on Consumer Electronics, IEEE, September 1989.

The Back Seat Driver: Real Time Spoken Driving Instructions (with J. Davis), Proceedings, IEEE Vehicle Navigation and Information Systems Conference, IEEE, Toronto, Canada, September 1989.

An Audio and Telephone Server for Multi-media Workstations (with M. McKenna), Proceedings, Second IEEE Conference on Workstations, IEEE, Palo Alto, CA., 1988.

Employing Voice Back Channels to Facilitate Audio Document Retrieval, Proceedings, ACM Conference on Office Information Systems (COIS), Santa Clara, CA, 1988.

Conversational Telecommunications Environments, Proceedings, Second International Conference on Human-Computer Interaction, 1987.

Understanding Speech Without Recognizing Words, Proceedings, American Voice Input/Output Society Conference, AVIOS, 1987.

A Robust Parser and Dialog Generator for a Conversational Office System (with B. Arons and C. Simmons), Proceedings, American Voice Input/Output Society Conference, AVIOS, Palo Alto, CA, 1987.

Integrated Messages and Network Services for a Personal Workstation, IEEE Workshop on Telematics and Message Handling Systems, IEEE, 1986.

Voice Interaction in an Integrated Office and Telecommunications Environment, Proceedings, American Voice Input/Output Society Conference, AVIOS, San Francisco, CA, 1985.

Voice Communication with Computers, book chapter in Advances in Human-Computer Interaction, H. R. Hartson ed., 1985.

Speech Synthesis Gives Voiced Access to an Electronic Mail System, Speech Technology, Vol. 2, No. 3, Aug/Sept 1984.

A Conversational Telephone Messaging System (with B. Arons), IEEE Transactions on Consumer Electronics, IEEE, Vol CE-30, August 1984.

Phone Slave: A Graphical Telecommunications Interface (with B. Arons), Proceedings, Society for Information Display International Symposium, SID, San Francisco, CA, June 1984.

Input/Display Registration in a Stereoscopic Workstation, Displays, April 1984.

Remote Access to Voice and Text Messages, Proceedings, American Voice Input/Output Society Conference, AVIOS, Washington D.C., 1984.

Fuzzy Fonts: Analog Models Improve Digital Text Quality, Proceedings, National Computer Graphics Association Conference, Chicago, IL, 1983.

Greyscale Fonts Designed From Video Signal Analysis, Society of Applied Learning Technology, Houston, TX, 1983.

Spatial Input/Display Correspondence in a Stereoscopic Computer Graphic Work Station, Proceedings, ACM/SIGGRAPH, Detroit, MI, 1983.

A Programmable Virtual Vocabulary Speech Processing Peripheral (with W. Bender), Proceedings, American Voice Input/Output Society Conference on Voice Data Entry Systems Applications, AVIOS, 1983.

The Intelligent Voice Interactive Interface (with E. A. Hulteen), Proceedings, Human Factors in Computer Systems, National Bureau of Standards/ACM, Gaithersburg, MD, 1982.

Interactive Three-Dimensional Computer Space, Proceedings, SPIE Conference on Processing and Display of Three-Dimensional Data, SPIE, San Diego, CA, 1982, Vol. 367.

Speech Communications, a Systems' Approach, Proceedings, American Voice Input/Output Society Conference on Entry Systems Applications, 1982.

Voice Interaction: Putting Intelligence into the Interface, Proceedings, IEEE International Conference on Cybernetics and Society, IEEE, Seattle, WA, 1982.

The Intelligent Ear: A Graphical Interface to Digital Audio, Proceedings, IEEE International Conference on Cybernetics and Society, IEEE, Atlanta, GA, 1981.

Soft Typography, Information Processing 1980, IFIPS, S. Lavington ed., North-Holland Publishing Co., 1980.
