

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner

v.

DIALECT, LLC,
Patent Owner

Case IPR2024-00749
U.S. Patent No. 7,693,720 B2
Issue Date: April 6, 2010

Title: MOBILE 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. 7,693,720 B2

TABLE OF CONTENTS

	Page
I. INTRODUCTION AND QUALIFICATIONS	12
A. Qualifications and Experience	12
B. Materials Considered	16
II. PERSON OF ORDINARY SKILL IN THE ART	18
III. STATEMENT OF LEGAL PRINCIPLES	20
A. Claim Construction	20
B. Obviousness (§ 103)	22
IV. THE '720 PATENT	28
A. Overview of the Specification	28
B. The Challenged Claims	31
V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIMS	33
A. Brief Summary of Prior Art	34
1. Prior Art Common to All Grounds	34
(a) Coffman (EX1003)	34
(b) Kanevsky (EX1004)	39
(c) Kneser (EX1005)	41
(d) Zadrozny (EX1006)	43
2. Additional Prior Art for Certain Dependent Claims	46
(a) DeLine (EX1010)	46
(b) Gerson (EX1011)	50
B. Ground 1: Claims 1-16, 22-25, 27, 29-31, 35-43, 52 and 55 Are Obvious Over Coffman (EX1003) in View of Kanevsky (EX1004), Kneser (EX1005), and Zadrozny (EX1006)	51
1. Claim 1	51

TABLE OF CONTENTS
(continued)

	Page
(a) “a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user and converts the received natural language speech utterance into an electronic signal; and” (1[a]).....	54
(b) “a natural language speech processing system connected to the computer device on the vehicle, wherein the natural language speech processing system receives, processes, and responds to the electronic signal using data received from a plurality of domain agents, wherein the natural language speech processing system includes:” (1[b])	63
(i) “a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;” (1[b][i]).....	75
(ii) “a parser that interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases, wherein the parser interprets the recognized words or phrases y:” (1[b][ii]).....	87

TABLE OF CONTENTS
(continued)

Page

(iii)	“an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.” (1[b][iii]).....	101
2.	Claim 2: “The mobile system according to claim 1, wherein the natural language speech processing system further includes an event manager that:” (2[pre]).....	112
(a)	“sends and receives events to a plurality of components of the natural language speech processing system to coordinate interaction among the plurality of components of the natural language speech processing system; and” (2[a])	113
(b)	“provides a multi-threaded environment that enables the natural language speech processing system to provide a plurality of real-time responses to a plurality of questions or commands in a plurality of different contexts across a plurality of overlapping or interleaved sessions with the user.” (2[b])	116
3.	Claim 3: “The mobile system according to claim 2, wherein the multi-threaded environment of the event manager further enables the natural language speech processing system-to use a plurality of the domain agents to provide the plurality of real-time responses to the plurality of questions or commands in the plurality of different contexts across the plurality of overlapping or interleaved sessions with the user.”.....	121

TABLE OF CONTENTS
(continued)

	Page
4. Claim 4: “The mobile system according to claim 1, wherein the response includes a text string and the natural language speech processing system further includes a text to speech engine that creates an encoded speech message to be annunciated to the user.”.....	122
5. Claim 5: “The mobile system according to claim 1, wherein the selected domain agent includes data associated with at least one of driving directions, travel information, restaurant information, systems information for the vehicle, safety information, or entertainment information.”	124
6. Claim 6: “The mobile system according to claim 1, wherein the selected domain agent includes data for communicating with one or more devices.”	125
7. Claim 7: “The mobile system according to claim 6, wherein the data for communicating with the one or more devices includes data for controlling the one or more devices.”	128
8. Claim 8: “The mobile system according to claim 6, wherein at least one of the one or more devices is connected to the vehicle.”	129
9. Claim 9: “The mobile system according to claim 8, wherein at least one of the speech unit or the natural language speech processing system is located remotely from the vehicle and connected to the computer device on the vehicle through a communication link.”	131
10. Claim 10: “The mobile system according to claim 8, wherein the at least one device connected to the vehicle includes at least one of a navigation system, a monitoring system, a security system, a control system, or a media system connected to the vehicle.”	135

TABLE OF CONTENTS
(continued)

	Page
11. Claim 11: “The mobile system according to claim 8, further comprising a telematics control unit that interfaces with the at least one device connected to the vehicle.”.....	135
12. Claim 12: “The mobile system according to claim 11, wherein at least one of the speech unit or the natural language speech processing system is embedded in the telematics control unit.”	138
13. Claim 13: “The mobile system according to claim 1, wherein the communicatively coupled services include at least one remotely located service and the selected domain agent includes data for controlling or communicating with the remotely located service.”	139
14. Claim 14: “The mobile system according to claim 13, wherein the remotely located service includes at least one of a payment service provider, a customer relationship management system, a specialized service, a location service, or an emergency service.”	140
15. Claim 15: “The mobile system according to claim 13, wherein the speech unit and the natural language speech processing system each include respective transceivers that communicate via a communication network.”	140
16. Claim 16: “The mobile system according to claim 15, wherein the communication network is a wide area wireless network.”	142
17. Claim 22: “The mobile system according to claim 1, wherein the speech unit is located remotely from the natural language speech processing system.”	143
18. Claim 23: “The mobile system according to claim 1, further comprising at least one of a display or a keypad.”	144

TABLE OF CONTENTS
(continued)

	Page
19. Claim 24: “The mobile system according to claim 1, wherein at least one of the speech unit or the natural language speech processing system is embedded in at least one of the vehicle, a handheld device connected to the vehicle, a fixed computer device connected to the vehicle, or a mobile computer device connected to the vehicle.”.....	145
20. Claim 25: “The mobile system according to claim 1, wherein the communicatively coupled services include at least one shared network resource.”	145
21. Claim 27: “The mobile system according to claim 1, wherein the one or more utilities included in the agent library include at least one of a text and string handling, function, a network communications function, a database lookup and management function, a fuzzy and probabilistic evaluation function, or a text to speech formatting function.”.....	146
22. Claim 29: “The mobile system according to claim 1, wherein the speech recognition engine:” (29[pre]).....	148
(a) “determines an identity of the user based on unique voice characteristics of the user; and” (29[a])	149
(b) “tags the recognized words or phrases with the identity of the user to associate the utterance with the user and the current dialog in the natural language speech processing system.” (29[b]).....	150
23. Claim 30: “The mobile, system according to claim 1, wherein the plurality of dictionary and phrase entries used by the speech recognition engine are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the one or more prior dialogs.”.....	152

TABLE OF CONTENTS
(continued)

	Page
24. Independent Claim 31: “A method responsive to a user generated natural language speech utterance, comprising:” (31[pre])	153
(a) “receiving, at a speech unit connected to a computer device on a vehicle, a natural language speech utterance from a user, wherein the speech unit converts the received natural language speech utterance into an electronic signal;” (31[a]).....	153
(b) “recognizing, at a speech recognition engine connected to the computer device on the vehicle, at least one of words or phrases from the electronic signal, wherein the speech recognition engine uses at least data received from a plurality of domain agents to recognize the words or phrases, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;” (31[b]).....	154
(c) “determining, at a parser connected to the computer device on the vehicle, a context for the natural language speech utterance;” (31[c])	155
(d) “selecting, at the parser connected to the computer device on the vehicle, at least one of the plurality of domain agents based on the determined context;” (31[d])	155
(e) “transforming, at the parser connected to the computer device on the vehicle, the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command; and” (31[e]).....	156

TABLE OF CONTENTS
(continued)

	Page
(f) “forwarding the formulated question or command to an agent architecture connected to the computer device on the vehicle, wherein the agent architecture communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.” (31[f]).....	157
25. Claim 35: “The method according to claim 31, the wherein the selected domain, [sic] agent includes data for controlling or communicating with at least one of a navigation system, a monitoring system, a security system, a control system, or a media system connected to the vehicle.”	158
26. Claim 36: “The method according to claim 31, wherein the selected domain agent includes data associated with at least one of driving directions, travel information, restaurant information, systems information for the vehicle, safety information, or entertainment information.”	159
27. Claim 37: “The method according to claim 31, wherein the communicatively coupled services include at least one remotely located service and the selected domain agent includes data for controlling or communicating with the remotely located service.”	159
28. Claim 38: “The method according to claim 37, wherein forwarding the formulated question or command to the agent architecture includes transmitting a request to the remotely located service.”	160

TABLE OF CONTENTS

(continued)

	Page
29. Claim 39: “The method according to claim 38, wherein the remotely located service is associated with a remotely located device.”.....	161
30. Claim 40: “The method according to claim 38, wherein the request is transmitted to the remotely located service via a communication network.”.....	162
31. Claim 41: “The method according to claim 37, wherein the remotely located service includes at least one of a payment service provider, a customer relationship management system, a specialized service, a location service, or an emergency service.”	162
32. Claim 42: “The method according to 38, wherein the request is transmitted to the remotely located service via a wide-area RF transceiver.”	163
33. Claim 43: “The method according to claim 31, wherein the communicatively coupled services include one or more shared network resources.”	164
34. Claim 52: “The method according to claim 31, further comprising providing interactive guidance to assist the user in performing one or more maneuvers to operate the vehicle.”.....	164
35. Claim 55: “The method according to claim 31, further comprising providing interactive guidance to relating to one or more tasks for the user based on a location of the vehicle.”.....	165
C. Ground 2: Claims 18, 20-21 and 32-33 Are Obvious Over Ground 1 Prior Art in Further View of DeLine (EX1010).....	166
1. Claim 18: “The mobile system according to claim 1, wherein the speech unit includes a speech coder that converts the natural language speech utterance into the electronic signal, an array microphone that receives the natural language speech utterance, and a filter that optimizes a signal to noise ratio of the electronic signal.”	166

TABLE OF CONTENTS
(continued)

	Page
2. Claim 20: “The mobile system according to claim 18, wherein the array microphone is at least a one-dimensional array.”.....	175
3. Claim 21: “The mobile system according to claim 18, wherein the speech coder uses an adaptive lossy audio compression to convert the natural language utterance into the electronic signal.”.....	176
4. Claim 32: “The method according to claim 31, wherein the speech unit includes an array microphone that receives the natural language speech utterance, a speech coder that converts the natural language speech utterance into the electronic signal, and a filter that optimizes a signal to noise ratio of the electronic signal.”.....	178
5. Claim 33: “The method according to 32, wherein the filter removes background noise from the electronic signal to optimize the signal to noise ratio of the electronic signal.”	178
D. Ground 3: Claims 19 and 34 Are Obvious Over Ground 2 Prior Art in Further View of Gerson (EX1011).....	179
6. Claim 19: “The mobile system according to claim 18, wherein the filter employs adaptive echo cancellation to optimize the signal to noise ratio of the electronic signal.”.....	179
7. Claim 34: “The method according to 32, wherein the filter employs at least one of adaptive echo cancellation or adaptive lossy audio compression to optimize the signal to noise ratio of the electronic signal.”.....	184
VI. NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS	184
VII. CONCLUSION	185

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 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. 7,693,720 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. 7,693,720 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. 7,693,720 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. 7,693,720 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. 7,693,720 B2 (“720 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. 7,693,720 B2

Exhibit No.	Description of Document
1001	U.S. Patent No. 7,693,720 B2 to Robert A. Kennewick et al. (filed July 15, 2003, issued April 6, 2010) (“ 720 ” or “ 720 patent ”)
1003	International Patent Application Publication No. WO 00/20962 to Daniel Coffman, Stephane H. Maes et al. (filed October 1, 1999, published April 13, 2000) (“ Coffman ”) ¹
1004	U.S. Patent No. 5,897,616 to Dimitri Kanevsky and Stephane Herman Maes (filed June 11, 1997, issued April 27, 1999) (“ Kanevsky ”)
1005	U.S. Patent No. 6,157,912 to Reinhard Kneser et al. (filed March 2, 1998, issued December 5, 2000) (“ Kneser ”)
1006	Wlodek Zadrozny et al., “Natural Language Understanding with a Grammar of Constructions”, Proc. 15th Int’l Conf. on Computational Linguistics (Vol. 2) (1994) (“ Zadrozny ”)
1007	International Patent Application Publication No. WO 00/21075 to Stephane H. Maes et al. (filed October 1, 1999, published April 13, 2000) (“ Maes ”) ²
1008	Excerpts from Dafydd Gibbon et al., <i>Handbook of Multimodal and Spoken Dialogue Systems : Resources, Terminology, and Product Evaluation</i> (2000) (“ Gibbon ”)
1009	Excerpts from <i>Microsoft Computer Dictionary</i> (5th ed. 2002)
1010	U.S. Patent No. 6,420,975 B1 to Jonathan E. DeLine et al. (filed Dec. 17, 1999, issued July 16, 2002) (“ DeLine ”)
1011	U.S. Patent No. 6,937,977 B2 to Ira A. Gerson (filed Oct. 5, 1999, issued August 30, 2005) (“ Gerson ”)

¹ I understand that Coffman also issued as U.S. Patent No. 7,137,126.

² I understand that Maes also issued as U.S. Patent No. 7,003,463.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

Exhibit No.	Description of Document
1012	Excerpts from <i>Digital Signal Processing in Telecommunications, European Project COST#229 Technical Contributions</i> , Aníbal R. Figueras-Vidal (Ed.) (1996)
1013	U.S. Patent No. 5,561,668 to Roland E. Genter (filed July 6, 1995, issued October 1, 1996)
1014	U.S. Patent Application Publication No. 2003/0144846 to Lawrence A. Denenberg and Christopher M. Schmandt (filed January 31, 2002, published July 31, 2003) (“ Denenberg ”)
1015	Excerpts from <i>Merriam-Webster’s Collegiate Dictionary</i> (10th ed. 1994)
1016	Excerpts from Richard F. E. Sutcliffe et al., <i>Industrial Parsing of Software Manuals</i> (1996) (“ Sutcliffe ”)
1017	Excerpts from <i>Newton’s Telecom Dictionary</i> (17th ed. 2001)
1019	’720 patent prosecution history

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 July 15, 2002, the filing date of the earliest patent application to which the

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

'720 patent claims priority (provisional appl. ser. no. 60/395,615).³

12. I have also been advised that to determine the appropriate level of a 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 '720 patent states that it "relates to the retrieval of online information and processing of commands through a speech interface in a vehicle environment." ('720, 1:12-14.) 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

³ I express no opinion on whether the '720 patent is actually entitled to the benefit of the provisional application to which it claims priority (60/395,615). This issue is not relevant here because all of the prior art references relied upon predate that date. In the event entitlement to the provisional filing date may later become relevant, for example based on arguments later presented by Patent Owner, I reserve my right to provide opinions on that issue at that time.

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 '720 patent have been based on the perspective of a person of ordinary skill in the art as of July 2002.

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

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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 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

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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 “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.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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.

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

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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 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

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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;

(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 '720 PATENT

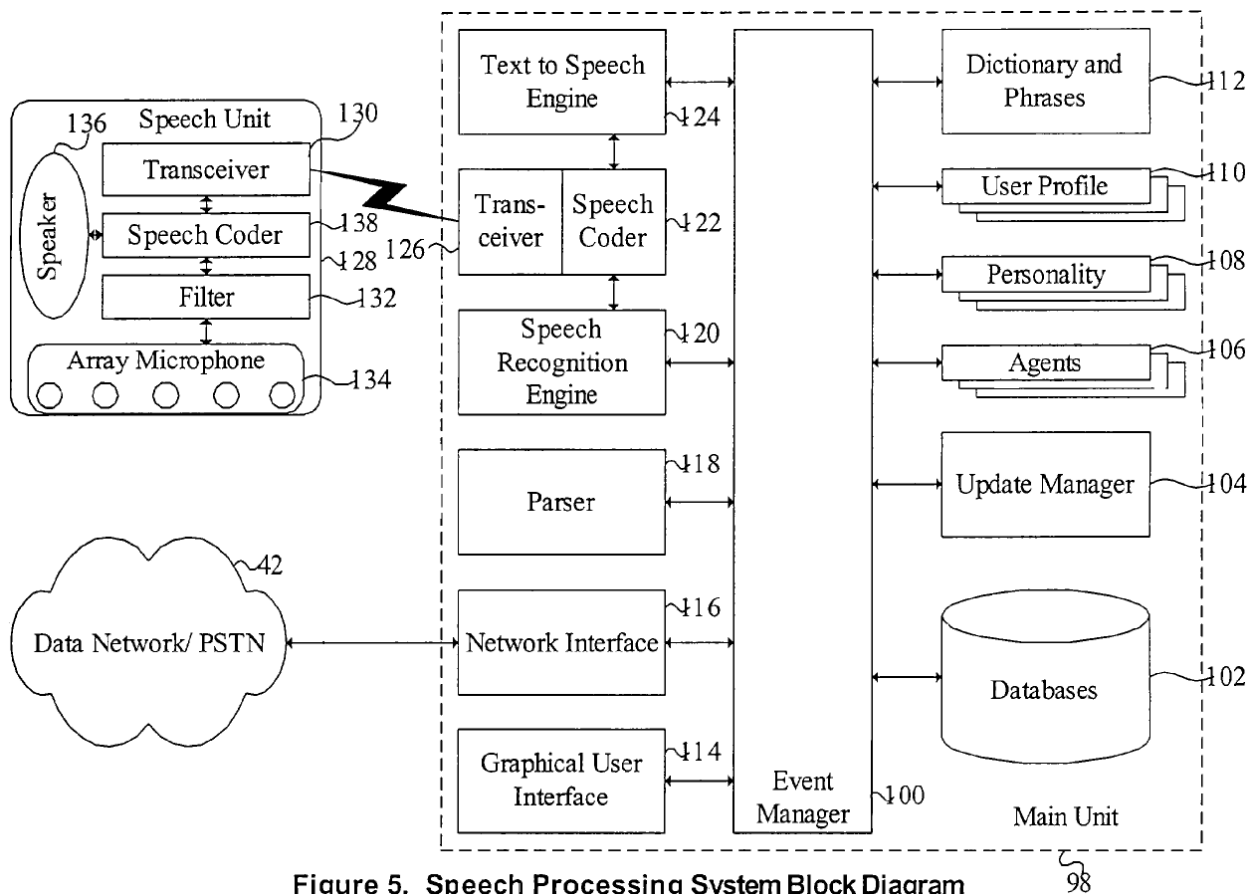
A. Overview of the Specification

32. The '720 patent states that “[t]he present invention relates to the retrieval of online information or processing of commands through a speech interface in a vehicle environment.” ('720, 1:12-14.)

33. The '720 patent admits that “[m]uch work covering multiple methods has been done in the fields of natural language processing and speech recognition,” 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 previously been applied to the parsing of speech queries.” ('720, 1:56-61.) According to the '720 patent, however, “no system developed provides a complete environment for users to make natural language speech queries or commands and receive natural sounding responses in a vehicular environment.” ('720, 1:61-64.) The '720 patent purports to “overcome...drawbacks of prior

speech-based telematic systems.” (’720, 2:48-49.)

34. Figure 5 of the ’720 patent, reproduced below, “is an overall diagrammatic view of the; [sic] interactive natural language speech processing system according to an embodiment of the invention” (’720, 15:44-46):



(’720, Fig. 5.) As shown in the figure above, the system includes, among other things, speech unit 128, speech recognition engine 120, dictionary and phrases 112, parser 118, and agents 106. Speech unit 128 includes a microphone to receive a spoken utterance from a user. (’720, 20:28-30.) After initial processing, the speech

is passed to speech recognition engine **120** for processing “using data in the dictionary and phrases module **112** and received from the agents **106**.” (’720, 20:64-66.) Then, “[t]he recognized words and phrases may be processed by the parser **118**, which transforms them into complete commands and questions using data supplied by the agents.” (’720, 20:67-21:2.) “The agents can then process the commands or questions.” (’720, 21:2-3.) In order to select the correct agent to process the command or question, “[t]he parser may determine a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile **110**, dialog history, and context stack contents.” (’720, 27:60-63.) “For example, a question with the keywords ‘temperature’ implies a context value of weather for the question.” (’720, 27:65-66.) “Once the context for the question or command has been determined, the parser **118** can invoke the correct agent **156**, **150**.” (’720, 28:25-27.) The agent returns results to the user. (’720, 21:11-12.)

35. With respect to “agents,” the ’720 patent states:

Agents are executables that receive, process and respond to user questions, queries and commands. The agents provide convenient and re-distributable packages or modules of functionality, typically for a specific domain. Agents can be packages of executable code, scripts, links to information, data, and other data forms, required to provide a specific/package of functionality, usually in a specific domain. In other words, an agent may include everything that is needed to extend the

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

functionality of the invention to a new domain. Further, agents and their associated data can be updated remotely over a network as new behavior is added or new information becomes available. Agents can use system resources and the services of other, typically more specialized, agents. Agents can be distributed and redistributed in a number of ways including on removable storage media, transfer over networks or attached to emails and other messages. An update manger is used to add new agents to the system or update existing agents.

(’720, 4:7-24.)

36. I discuss additional aspects of the ’720 patent as needed in the analysis below.

B. The Challenged Claims

37. This Declaration addresses claims 1-16, 18-25, 27, 29-43, 52 and 55 of the ’720 patent. Independent claim 1 recites:

1. A mobile system responsive to a user generated natural language speech utterance, comprising:
 - [a] a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user and converts the received natural language speech utterance into an electronic signal; and
 - [b] a natural language speech processing system connected to the computer device on the vehicle, wherein the natural language speech processing

system receives, processes, and responds to the electronic signal using data received from a plurality of domain agents, wherein the natural language speech processing system includes:

- [c] a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;
- [d] a parser that interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases, wherein the parser interprets the recognized words or phrases by:
 - [d][i] determining a context for the natural language speech utterance;
 - [d][ii] selecting at least one of the plurality of domain agents based on the determined context; and
 - [d][iii] transforming the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command; and
- [e] an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and

an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.

(’720, 32:30-33:6 (Claim 1) (bracketed notation (e.g., “[a]”) added).)

38. I address the claims further in my detailed analysis in **Part V** below.

V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIMS

39. I have reviewed and analyzed the prior art references and materials listed in **Part I.B** above. In my opinion the claims of the ’720 patent would have been obvious to a person of ordinary skill in the art based on the following combinations of the prior art.

Ground	Claims	Basis for Challenge under § 103
1	1-16, 22-25, 27, 29-31, 35-43, 52, 55	Coffman (EX1003), Kanevsky (EX1004), Kneser (EX1005), and Zadrozny (EX1006)
2	18, 20-21, 32-33	Ground 1 Prior Art + DeLine (EX1010)
3	19, 34	Ground 2 Prior Art + Gerson (EX1011)

40. I am informed by counsel that each of the references cited in the grounds above qualifies as prior art to the challenged claims because each reference was filed and/or published before the ’720 patent priority date (July 15, 2002).

A. Brief Summary of Prior Art

1. Prior Art Common to All Grounds

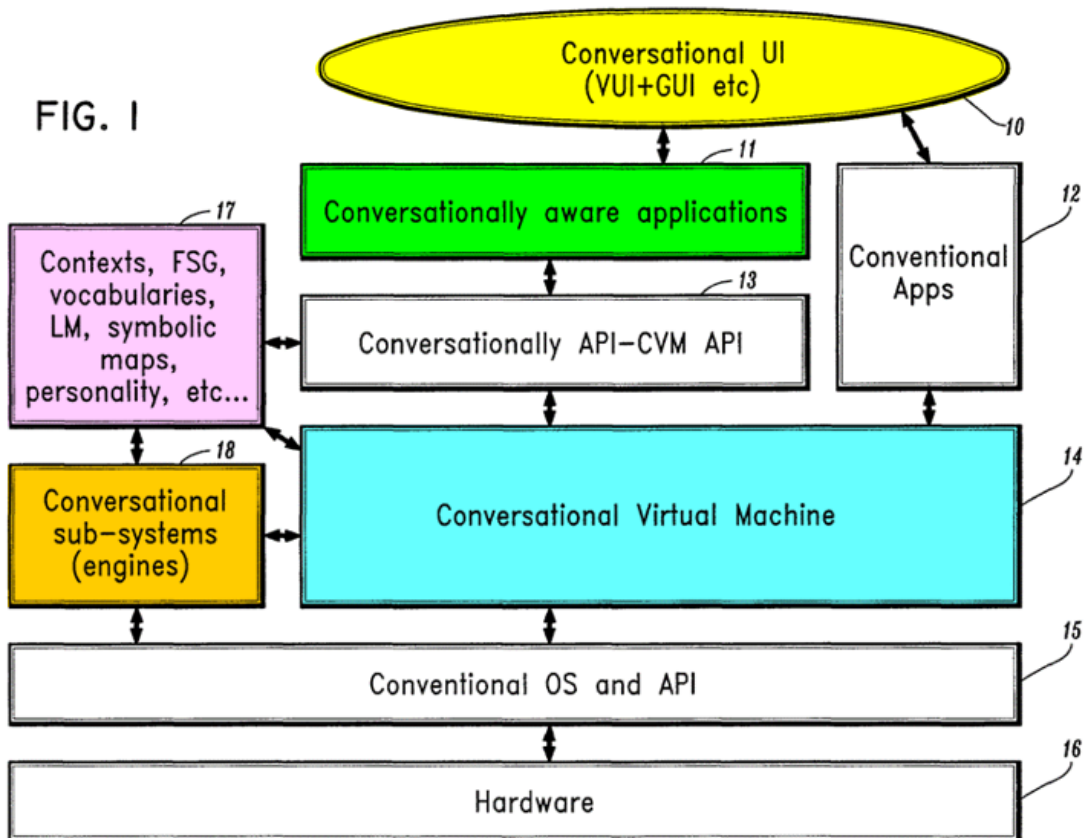
(a) Coffman (EX1003)

41. Coffman, International Patent Application Publication No. WO 00/20962 to Daniel Coffman et al. is entitled “Conversational Computing via Conversational Virtual Machine.” The face page of Coffman identifies International Business Machines as the applicant. (Coffman, face page (field 71).) I am informed that Coffman qualifies as prior art because it published on April 13, 2000, before the earliest priority date of the ’720 patent.

42. Coffman is a strong reference that discloses or suggests the majority of the limitations of the challenged claims. And as explained below, it also incorporates by reference Kanevsky [EX1004] (and Maes [EX1007]), which in turn cites Zadrozny [EX1006].

43. Coffman describes a “conversational computing system.” Generally speaking, a conversational system is a computing system with either purely voice input/output or a multi-modal system with voice input/output (as well as other input/output means). (Maes, 1:12-13.) Coffman’s system “provides the capability to have a natural dialog with NLU [natural language understanding], NLG [natural language generation] and mixed initiative across multiple applications, with multiple

domains.” (Coffman, 36:29-31.) Figure 1 of Coffman provides a high-level depiction of the system’s components:



(Coffman, Fig. 1 (highlighting added).) As shown above, the system includes, among other things, “a universal coordinated multi-modal **conversational user interface (CUI) (10)**,” “a plurality of **con conversationally-aware applications (11)**,” “**con conversational kernel (14)**,” “**con conversational arguments (17)**,” and “**con conversational engines (18)**.” (Coffman, face page (Abstract).)

44. The **con conversational virtual machine (or CVM)** is “[a] key component” of the conversational system that “allows a computer or any other interactive device

to converse with a user.” (Coffman, 10:23-25, 11:10-13.) To do this, the CVM uses **conversational subsystems** “including speech recognition, speaker recognition, text-to-speech, natural language understanding and natural dialog generation engines to understand and generate dialog between user and machine.” (Coffman, 5:1-4.) The **conversational subsystems** “are responsible for converting voice requests into queries and converting outputs and results into spoken messages using the appropriate **data files** 17 (e.g., contexts, finite state grammars, vocabularies, language models, symbolic query maps etc.)” (Coffman, 13:8-11.)

45. “The **conversationally aware applications** 11 are applications that are specifically programmed for operating with a **CVM core layer (or kernel)** 14 via conversational application APIs 13.” (Coffman, 12:28-30.) A **conversationally aware application** operates with a **CVM** by providing its **data files** to the CVM for use by the **conversational subsystems**—an application hooks into the CVM API by “registering all its conversational capabilities and requirements” including “[a]ctive vocabulary, finite state grammar and language models to control the application” and “[s]ymbolic mapping if NLU is supported or at list [sic; least] a context state list.” (Coffman, 57:23-27; *see also id.*, 36:31-32 (“each application will provide the CVM 401 its own parsing and translation arguments”).)

46. The CVM also “allows the user to run multiple tasks on a machine.”

(Coffman, 11:10-13.) The CVM is able to do this by using a “context stack,” which enables the system to determine the appropriate conversationally-aware application for handling user input. (Coffman, 42:16-17 (“This stack structure allows to maintain non-ambiguous conversational interactions with multiple tasks, threads or processes.”).) The context stack “comprises all the information associated with an application.” (Coffman, 21:23-24.) “Such information includes all the variable, states, input, output and queries to the backend that are performed in the context of the dialog and any extraneous event that occurs during the dialog.” (Coffman, 21:24-26.) Coffman explains:

In general, the context stack 405 accumulates the context (i.e., full query arguments list of attribute value n-uples, and state/mode) of each active process with an activated discourse (i.e., conversational interaction associated with a given task/process/thread) along with any data files 413 (or at least identifiers of such conversational arguments) for the different engines that may be needed for input understanding (e.g., files or arguments that the engines use for performing their respective tasks such as active FSG, topic, vocabulary, HMM (hidden markov models), voiceprints, language models or possible queries for a speech input). In other words, the term “context” refers to the state of each discourse (whether active or nonnative), which keeps track of the past history of the discourse, its current state, and the specific characteristics and full query arguments of the corresponding task (e.g,

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

vocabulary file, language model, parsing, tags, voiceprint, TTS rules, grammar, NLU etc. of each active task/process) along with any data needed for input understanding.

(Coffman, 29:20-32.) To identify the appropriate conversationally-aware application to handle a user's input (such as a user's spoken utterance), the system searches the context stack for a context that is an acceptable match:

Completion/search for the active context is performed from context to context down the stack. That is, new queries or arguments are compared by the dialog engine by going down the stack until an acceptable match is obtained and optionally confirmation is obtained from the user. As soon a context is found that fits the utterance at the level of the NLU symbolic language, the context becomes active and the corresponding process becomes active.

(Coffman, 42:1-5.) "The first task to have its dialog completed by producing a non-ambiguous query is executed and the corresponding query as interpreted by the other application is stored to activate if the recognized query is rejected by the user."

(Coffman, 37:2-4; *see also id.*, 42:18-19 ("If the request is complete, it will be executed, pending possible request for confirmation by the user, e.g. when it is irreversible.").)

47. I discuss additional aspects of Coffman as needed in my analysis of the challenged claims below.

(b) Kanevsky (EX1004)

48. Kanevsky, U.S. Patent No. 5,897,616 to Dimitri Kanevsky and Stephane Herman Maes is entitled “Apparatus and Methods for Speaker Verification/Identification/Classification Employing Non-Acoustic and/or Acoustic Models and Databases.” I am informed that Kanevsky qualifies as prior art because it was filed on June 11, 1997 and issued on April 27, 1999, before the earliest priority date of the ’720 patent.

49. I note that Coffman expressly incorporates Kanevsky by reference. (Coffman, 37:10-13.) I am informed by counsel that under the patent laws, when a document is “incorporated by reference” into a host document, the incorporated document’s material effectively becomes part of the host document as if it were explicitly contained therein. Therefore, I am informed that Kanevsky is properly considered part of the disclosures of Coffman for purposes of applying it as prior art. I further note that both references also share a named co-inventor, Stephane Herman Maes, and identify IBM as the assignee/applicant. (Coffman, face page; Kanevsky, face page.)

50. My analysis points to Kanevsky primarily for Kanevsky’s further details about speech recognition. Claim 1, limitation 1**[b][i]**, of the ’720 patent recites “**data used by the speech recognition engine [that] includes a plurality of**

dictionary and phrase entries.” As mentioned above, Coffman describes a conversational computing system that uses speech recognition and also describes how a conversationally-aware application provides data files for controlling the application, including active vocabulary, finite state grammar and language models. Coffman describes using the language models as part of “converting voice requests into queries” (Coffman, 13:8-11) but does not appear to further explain what a language model is or how it is used for speech recognition. That is not surprising because language models and their use for speech recognition were well known. This is confirmed by Kanevsky, which explains that a speech recognition engine may be implemented using “Hidden Markov Models, mixtures of Gaussian probabilities, mel cepstral vectors as acoustic features, a 20K or 64K vocabulary and a trigram language model.” (Kanevsky, 13:50-56.) Kanevsky describes such a speech recognition engine as an “existing component” and explains that “[s]uch systems, for example, are disclosed in IBM’s speech dictation engine ‘Simply Speaking’ and in ‘Transcription of Radio Broadcast News with the IBM Large Vocabulary Speech Recognition System’, P.S. Gopialakrishnan [sic, Gopalakrishnan] et al., Proceeding Speech Recognition Workshop, Arden House, Drapa (1996).” (Kanevsky, 13:48-61.)

51. As I discuss in my analysis below, a “trigram language model” was a

well-known type of “n-gram language model” comprising sequences of three words (i.e., “trigrams”) and the probabilities of their occurrence based (at least initially) on some previously collected representative set of text, called training data. As I explain below, Kanevsky and its a trigram language model teachings disclose the claimed “**plurality of dictionary and phrase entries.**”

52. I discuss additional aspects of Kanevsky as needed in my analysis of the challenged claims below.

(c) Kneser (EX1005)

53. **Kneser**, U.S. Patent No. 6,157,912 to Reinhard Kneser et al. is entitled “Speech Recognition Method with Language Model Adaptation.” I am informed that Kneser qualifies as prior art because it was filed on March 2, 1998 and issued on December 5, 2000, before the earliest priority date of the ’720 patent.

54. My analysis points to Kneser for its teachings regarding adaptive language models. Claim 1, limitation 1**[b][i]**, of the ’720 patent recites “**a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user.**” As I noted above, a language model, such as a trigram language model disclosed by Kanevsky, comprises “**a plurality of dictionary and phrase entries.**” An “adaptive” language model is a language model that can change as the model is

used, such that it “adapts” to reflect that usage history. Although neither Coffman nor Kanevsky appears to specifically mention using them, adaptive n-gram language models and their advantages were well known, which Kneser confirms. Kneser describes updating a language model’s probabilities that “given word sequences of limited length” (i.e., n-grams) are present based on the user’s interaction with the system, explaining:

These language model values represent the probabilities that given word sequences of limited length are present; they have been obtained from a comparatively large quantity of training texts, which may also be formed by written texts. Long-term changes in the use of language, for example subject matter or style, can accordingly not be taken into account with these language model values [probabilities]. Since the quality of the adaptation of the language model values has a considerable influence on the reliability of the speech recognition, it is desirable to adapt the language model values as well as possible to the given text by means of a small quantity of updated text. This adaptation may take place during the process of recognizing a given text or alternatively subsequent to the recognition of part of a text, after recognition errors have been manually corrected.

(Kneser, 1:22-29; *see generally id.*, 1:13-37.)

55. As explained below, based on Kneser’s teachings, it would have been obvious for the language models described by the combination of Coffman and

Kanevsky to be “adaptive,” thereby disclosing a “**a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user.**”

56. I discuss additional aspects of Kneser as needed in my analysis of the challenged claims below.

(d) Zadrozny (EX1006)

57. **Zadrozny** is a paper authored by Wlodek Zadrozny et al. entitled “NL Understanding with a Grammar of Constructions” and published in volume 2 of the Proceedings of the 15th International Conference on Computational Linguistics in 1994. I am informed that Zadrozny qualifies as prior art because its 1994 publication date is before the earliest priority date of the ’720 patent.

58. In another link between prior art references, Zadrozny is cited in Kanevsky as an example of a semantic analyzer. (Kanevsky, 14:22-25.) Zadrozny also identifies a number of its co-authors (including Wlodek Zadrozny) as being associated with IBM Research Division. (Zadrozny, p.1.)

59. My analysis points to Zadrozny primarily for its teachings regarding using a grammar associated with an application to perform natural language understanding and formulate an application query. Claim 1, limitation 1 **[b][ii][3]**, of the ’720 patent recites “**at least one question or command...formulated in a**

grammar that the selected domain agent uses to process the formulated question or command. Zadrozny describes a “grammar of constructions” associated with an application (such as a calendar application) for performing natural language understanding to translate a recognized utterance into “a form that can be executed by the application”:

We present an approach to natural language understanding based on a computable grammar of constructions. A construction consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system. . . .

With respect to parsing, grammars of constructions can be parsed like “standard” grammars, except that the set of features is richer. Given a string (representing a sentence, a fragment of a discourse or a paragraph), the parser assigns it a construction.... Here the message of an utterance must be understood in the context of an intended action. This is done in two steps. First, the system determines the intended action and its parameters, using domain knowledge (meetings+time+places). Second, once all the parameters have been extracted from the dialog, the system executes the action. To do this, the program uses application-specific knowledge to translate the action and its parameters into a form that can be executed by the application (Xdiary).

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

(Zadrozny, pp.1-2.) As explained above in the excerpt above, Zadrozny describes a technique using a “grammar of constructions” to formulate an utterance as an intended action and parameters and then translate the action and parameters into a form for the application to execute. Zadrozny illustrates the technique with the example utterance “I want you to arrange a conference in my office at 5.” (Zadrozny, p.3.) Using a grammar of constructions, the utterance is “analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, verb, an object and a complement” and formulated as an action (“schedule”) and series of parameters:

```
[ [ action_name schedule]
  [ event_name
    [ a conference]
  [ event_time
    [ [ minute 0]
      [ hour
        [ 5 am_or_pm]
      ]
    ]
  [ event_place
    [ my office]
```

(Zadrozny, pp.3-4.) As Zadrozny explains, this action and parameters are then translated into a form for the application to execute (i.e., an application call/query).

60. As explained below, it would have been obvious to incorporate Zadrozny’s teachings regarding a grammar of constructions into Coffman’s conversational system in order to translate a user’s recognized spoken utterance into an application-specific query.

61. I discuss additional aspects of Zadrozny as needed in my analysis of the challenged claims below.

2. Additional Prior Art for Certain Dependent Claims

(a) DeLine (EX1010)

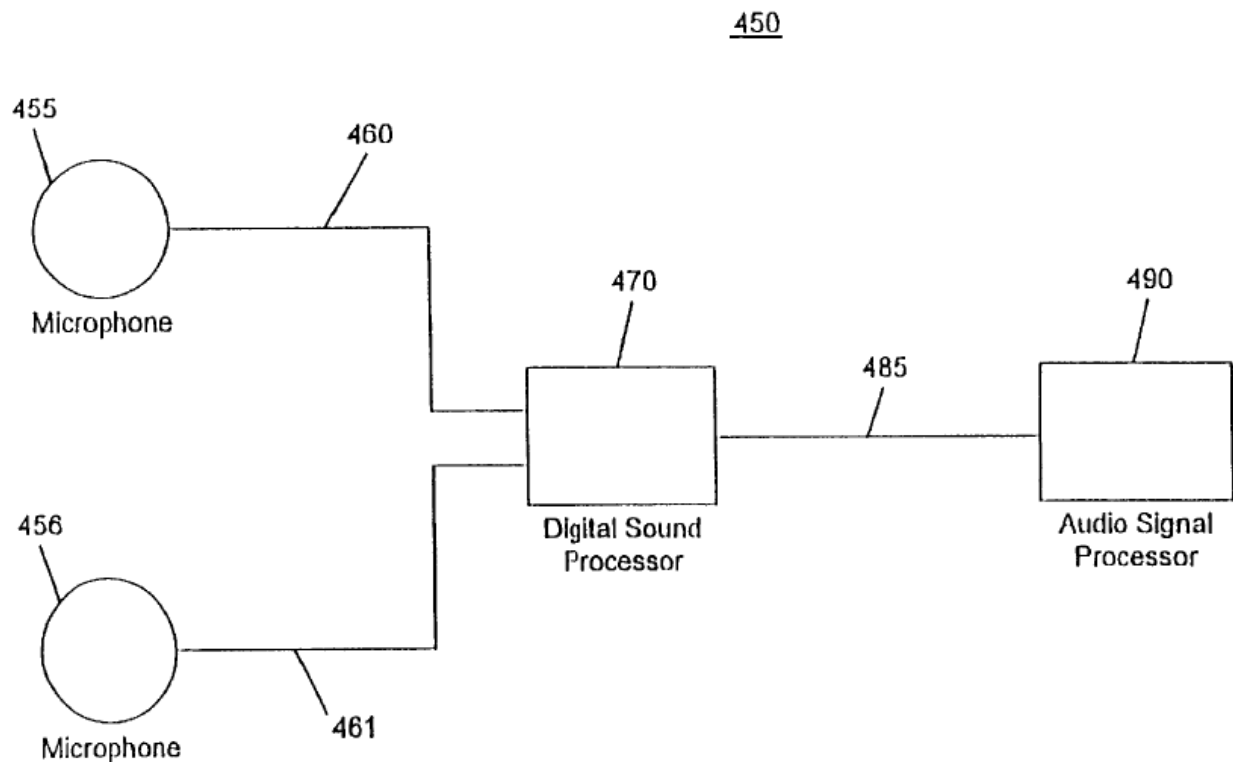
62. **DeLine**, U.S. Patent No. 6,420,975 B1 to Jonathan E. DeLine et al. is entitled “Interior Rearview Mirror Sound Processing System.” I am informed that DeLine qualifies as prior art because it was filed on December 17, 1999, before the earliest priority date of the ’720 patent.

63. My analysis points to DeLine in connection with limitations recited in dependent claims related to a speech unit including an “**array microphone**” and “**filter**” that optimizes the signal to noise ratio of an electronic signal.

64. DeLine generally describes an in-vehicle microphone arrangement and sound processing system for enhancing vocal content and reducing ambient noise content detected by the microphone(s). (DeLine, 1:20-28.) DeLine explains that many vehicles use hands-free communication devices to avoid problems that may arise when a driver has to hold the device while driving. (DeLine, 1:29-32.) But with such devices, DeLine explains, the vocal content detected by the microphone may be degraded. (DeLine, 1:40-2:33.) For example, “ambient noise from the like of traffic noise, road noise, wind noise, HVAC noise and engine noise can make it

difficult for the remote party to voice recognize with clarity and accuracy the message being telecommunicated.” (DeLine, 2:29-33.) DeLine explains that “[t]his may be especially troublesome when the other party is an automated voice system which responds to a particular voice communication.” (DeLine, 2:18-20.) “When a remote receiving system is operating in voice recognition mode, it is particularly important that alphanumeric speech text is accurately received and interpreted by the remote party.” (DeLine, 2:20-23; *see generally id.*, 2:15-33.)

65. In one embodiment of its system, DeLine proposes a microphone array comprising multiple microphones, as illustrated in Figure 12 below:



(DeLine, Fig. 12; *see also id.*, 47:19-53, Fig. 14 (depicting two-microphone mirror-

mounted system comprising microphones **520** and **530**), 48:53-63 (describing “array of multiple mirror-mounted microphones” and commercially-available microphone array from Andrea Electronics Corporation).) “FIG. 12 describes a vehicular digital sound processing **450** utilizing two microphones **455**, **456** that provide signals **460**, **461**, respectively, as inputs to digital sound processor **470** which provides vocal signature signal **485** to audio vehicle signal processor **490**.” (DeLine, 47:19-23.) DeLine explains the benefits and advantages of using input from multiple microphones:

By receiving separate inputs (and with each individual input optionally filtered to enhance their principal signal content), digital sound processor **470** can mathematically/computationally analyse the respective inputs, and their combination, in order to extract, using noise cancellation techniques as known in the art, vocal signature signal **485** which is input to audio vehicle signal processor **490** (for example, an ONSTAR™ system) where signal **485** is clearly and accurately received and processed, including in voice recognition mode. Use of at least two microphones in vehicular digital sound processing is beneficial to facilitate generation of high quality vocal signals because destructive interference techniques and/or signal averaging techniques can be utilized to cancel out unwanted noise and to isolate the desired vocal signature signal that is indicative of the alphanumeric words spoken by the driver and/or other vehicle occupant.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

(DeLine, 47:32-48.) As noted in the excerpt above, DeLine describes how the use of multiple microphones enables a digital sound processor to use “noise cancellation techniques as known in the art” to “facilitate generation of high quality vocal signals.” (*Id.*) “Such noise-cancelling techniques,” DeLine explains, “include techniques such as echo cancellation including echo cancellation algorithms; beam forming techniques including beamforming digital algorithms and adaptive beamforming; active noise cancellation; open and closed loop noise cancellation; adaptive signal processing techniques and adaptive noise cancellation techniques such that when, for example, speech by a driver is received at and detected by microphone **520**, the output from the opposing microphone **530** is analyzed for ambient noise content (and vice versa should speech by a passenger be detected by microphone **530**).” (DeLine, 48:26-45.)

66. As explained in my analysis below, DeLine’s use of two or more microphones discloses the claimed “**array microphone**.” And DeLine’s discussion of system components to perform various noise-cancelling techniques, discloses the claimed “**signal**.” (DeLine, 47:32-48, 48:26-45.)

67. I discuss additional aspects of DeLine as needed in my analysis of the challenged claims below.

(b) Gerson (EX1011)

68. **Gerson**, U.S. Patent No. 6,937,977 B2 to Ira A. Gerson is entitled “Method and Apparatus for Processing an Input Speech Signal During Presentation of an Output Audio Signal.” I am informed that Gerson qualifies as prior art because it was filed on October 5, 1999, before the earliest priority date of the ’720 patent.

69. My analysis points to Gerson in connection with limitations recited in dependent claims related to a speech unit including a “**filter**” that employs “**adaptive echo cancellation**.”

70. Gerson “relates generally to communication system incorporating speech recognition and, in particular, to a method and apparatus for ‘barge-in’ processing of an input speech signal during presentation of an output audio signal.” (Gerson, 1:8-12.) Gerson explains that “[g]iven human nature and the design of some speech synthesis/recognition systems, the spoken responses provided by a caller will often occur during presentation of an output audio signal, for example, a synthesized speech prompt.” (Gerson, 1:35-39.) “The processing of such occurrences is often referred to as ‘barge-in’ processing.” (Gerson, 1:39-40.) In a “barge-in” situation, “the speech recognition system must account for residual artifacts from the prompt being present in any spoken response provided by the user (i.e., an input speech signal) in order to effectively perform speech recognition.”

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

(Gerson, 1:44-50.)

71. Gerson's system includes an "echo cancellation and environmental processing (ECEP) block" that receives input from a microphone and "provides echo cancellation of the speaker audio **211** from the microphone audio **220**." (Gerson, 9:5-11, 9:23-26.) Gerson explains that "[t]his form of echo cancellation is known as acoustic echo cancellation and is well known in the art" and cites U.S. Patent No. 5,561,668 as an example (which it incorporates by reference). (Gerson, 9:26-34; *see generally id.*, 9:24-10:10 (describing echo cancellation).) As I explain in my analysis below, a person of ordinary skill in the art would have understood that Gerson echo cancellation techniques disclose "**adaptive echo cancellation.**"

72. I discuss additional aspects of Gerson as needed in my analysis of the challenged claims below.

B. Ground 1: Claims 1-16, 22-25, 27, 29-31, 35-43, 52 and 55 Are Obvious Over Coffman (EX1003) in View of Kanevsky (EX1004), Kneser (EX1005), and Zadrozny (EX1006).

1. Claim 1

73. For reference, claim 1 recites:

1. A mobile system responsive to a user generated natural language speech utterance, comprising:
 - [a]** a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user

and converts the received natural language speech utterance into an electronic signal; and

[b] a natural language speech processing system connected to the computer device on the vehicle, wherein the natural language speech processing system receives, processes, and responds to the electronic signal using data received from a plurality of domain agents, wherein the natural language speech processing system includes:

[b][i] a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;

[b][ii] a parser that interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases, wherein the parser interprets the recognized words or phrases by:

[b][ii][1] determining a context for the natural language speech utterance;

[b][ii][2] selecting at least one of the plurality of domain agents based on the determined context; and

[b][ii][3] transforming the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command; and

[b][iii] an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.

(’720, 32:30-33:6 (Claim 1) (bracketed notation (e.g., “[a]”) added).)

74. The preamble recites “[a] **mobile system responsive to a user generated natural language speech utterance, comprising.**” As discussed below, Coffman discloses a “**mobile system**” comprising a conversational computing system involving a “conversational car PC (PVA, Personal Vehicle Assistant)” that is “**responsive to a user generated natural language speech utterance.**” (Coffman, 56:21-25 (“[Y]ou have, in your car, a conversational PDA, a conversational car PC (PVA, Personal Vehicle Assistant) and a conversational smart phone. Further assume that the PVA runs an embedded CVM while the two other

applications are conversationally aware, i.e., you can control them through the CVM running on the PVA.”), 55:27-28 (“When a user interacts with the collection of devices, the interaction may always be via a central unit such as a PVA (personal vehicle assistant) 710 in a car, or a speech browser 713.”).⁴ As shown below, the claimed system is disclosed and rendered obvious by the prior art.

(a) **“a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user and converts the received natural language speech utterance into an electronic signal; and” (1[a])**

75. The written description of the '720 patent describes a “**speech unit**” as comprising a microphone and speaker, with other optional elements such as a filter, speech coder, and transceiver. ('720, 20:28-35 (“The speech unit **128** includes one or more microphones, for example array microphone **134**, to receive the utterances of the user.... The speech received at the microphone **134** may be processed by filters **132** and passed to the speech coder **138** for encoding and compression. In one embodiment, a transceiver module **130** transmits the coded speech to the main unit **98**.”), 20:35-39 (“Coded speech received from the main unit **98** is detected by the transceiver **130**, then decoded and decompressed by the speech coder **138** and

⁴ 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. 7,693,720 B2

announced by the speaker **136.**”), 20:52-55 (“In an alternative embodiment, wires connect the speech unit **128** and the main unit **98**, eliminating the need for one speech coder **138.**”), Fig. 5 (showing speech unit including, among other things, microphone **134** and speaker **136.**.) The speech unit and a main unit processing device can be integrated into a single physical device or connected via wires or wirelessly (e.g., transceiver). (’720, 18:62-64 (“In some embodiments, the main speech processing unit **98** and speech unit **128** may be embedded into the handheld computer **36.**”); *see also id.*, Fig. 3, 18:53-57 (“It will be understood that the exact distribution and packaging of the main speech processing unit **98** and speech unit **128**, can be determined by the details of the deployment situation and will not change the functionality, capabilities or spirit of the invention in any way.”), 3:1-5 (“The speech unit may be incorporated into a vehicle computer device or system, or may be a separate device. If a separate device, the speech unit may be connected to the vehicle computer device via a wired or wireless connection.”), 20:33-35, 20:52-55, Fig. 5.)

76. A person of ordinary skill in the art would have understood and found it obvious that Coffman’s system would comprise “**a speech unit connected to a computer device**” as claimed. Like the ’720 patent, Coffman similarly describes user devices for its conversational computing system having a microphone and speaker—i.e., a “**speech unit**”—for receiving speech input from a user and

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

providing audio output (e.g., via text-to-speech). (Coffman, e.g., 53:8-10 (“Voice input from a microphone (e.g. the standard voice input) can be arguments of function calls with the sequence of words, phones, or user identity or queries (symbolic language representation provided by NLU).”), 37:14-22 (“Spoken input from a speech client (e.g., telephone 600) is subject to a speech recognition process 604 and other input (e.g., keyboard, mouse clicks etc) are subject to NLU processing 605.”), 35:2-3 (“listen to the output of a task”), 39:12 (“audio only output”), 39:30 (“voice output”), Fig. 6 (telephone 600, “speech synthesis,” “speech out”).) Coffman further teaches that these components are coupled to components for processing input from and output to them (“**computer device**”), such as conventional processing components (e.g., CPU) of a desktop computer, PDA, or “conversational car PC (PVA, Personal Vehicle Assistant)” for running a conversational virtual machine (“CVM”) and related software. (Coffman, 10:26-28 (“A CVM platform can also be implemented with PvC (pervasive computing) clients as well as servers.”), 12:5-7 (“Referring now to Fig. 1, a block diagram illustrates a conversational computing system (or CVM system) according to an embodiment of the present invention, which may be implemented on a client device or a server.”), 13:1-3 (“The CVM system may be built on top of a conventional OS and APIs 15 and conventional device hardware 16 and located on a server or any client device (PC, PDA, PvC).”),

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

56:21-25 (“[Y]ou have, in your car, a conversational PDA, a conversational car PC (PVA, Personal Vehicle Assistant) and a conversational smart phone. Further assume that the PVA runs an embedded CVM while the two other applications are conversationally aware, i.e., you can control them through the CVM running on the PVA.”), 28:29-31 (“In general, the CVM 401 operates by converting conversational and conventional input streams into multiple actions and produces sorted output to a user through conversational and/or conventional output streams.”), 20:26-29 (describing prioritizing “CPU cycles”), 45:29-32 (“CPU load” and “CPU cycles”), Figs. 1 & 6.)

77. A person of ordinary skill in the art would have further understood and found it obvious that Coffman’s system would comprise **“a speech unit connected to a computer device on a vehicle.”** As noted above, Coffman explains that the computer device that receives speech input can take the form of a “conversational car PC (PVA, Personal Vehicle Assistant).” (Coffman, 56:21-25 (“[Y]ou have, in your car, a conversational PDA, a conversational car PC (PVA, Personal Vehicle Assistant) and a conversational smart phone. Further assume that the PVA runs an embedded CVM while the two other applications are conversationally aware, i.e., you can control them through the CVM running on the PVA.”), 55:27-28 (“When a user interacts with the collection of devices, the interaction may always be via a

central unit such as a PVA (personal vehicle assistant) 710 in a car, or a speech browser 713.”.)

78. Coffman further discloses **“the speech unit receives a...speech utterance from a user.”** Coffman describes receiving spoken input from a person using, for example, a telephone or microphone. (Coffman, e.g., 53:8-10 (“Voice input from a microphone (e.g. the standard voice input) can be arguments of function calls with the sequence of words, phones, or user identity or queries (symbolic language representation provided by NLU.”), 37:14-22 (“Spoken input from a speech client (e.g., telephone 600) is subject to a speech recognition process 604 and other input (e.g., keyboard, mouse clicks etc) are subject to NLU processing 605.”), 41:2-5 (“Based on the modality (pointer, keyboard, file, speech), the task dispatcher 402 redirects the stream to the appropriate conversational subsystems or conventional subsystem with speech inputs being transcribed/understood. The output of these subsystems is run down the context stack 405 to extract the active query and complete it.”), 53:22-26 (“The input process described above in accordance with one aspect of the present invention considers that speech or any other input stream is included as a classical input/output stream that is susceptible to all forms of processing typically reserved for character or binary input.”.)) Figure 6 of Coffman, reproduced below, illustrates receiving a user generated speech

utterance via telephone 600:

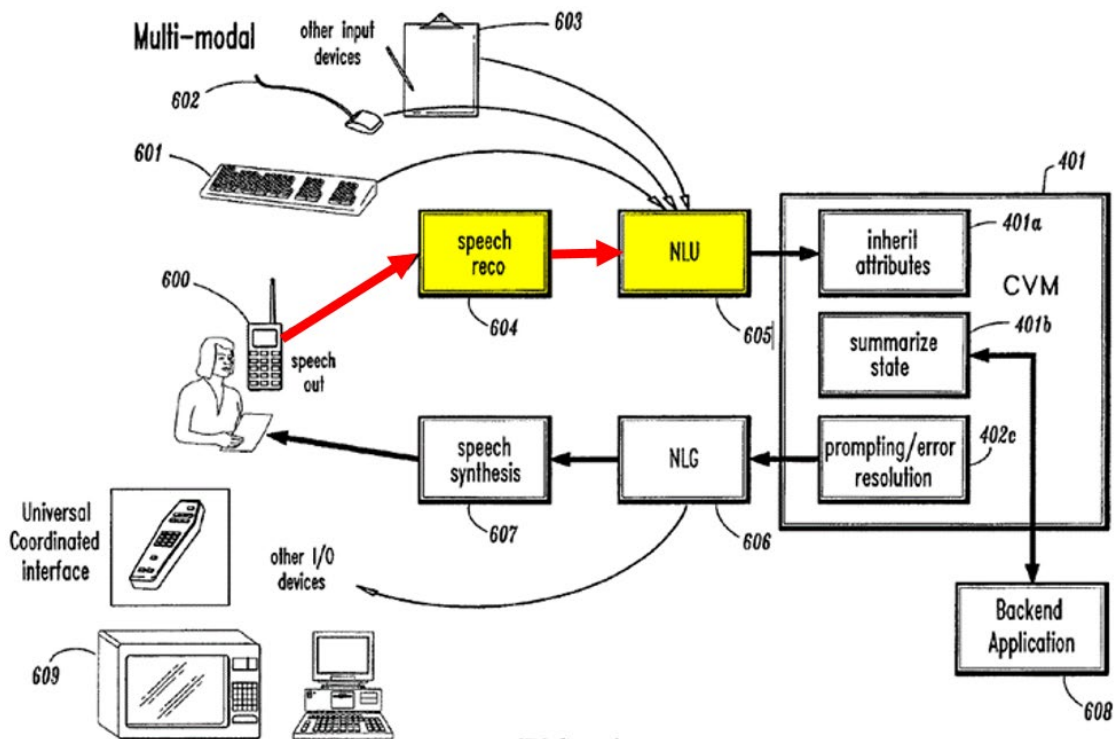


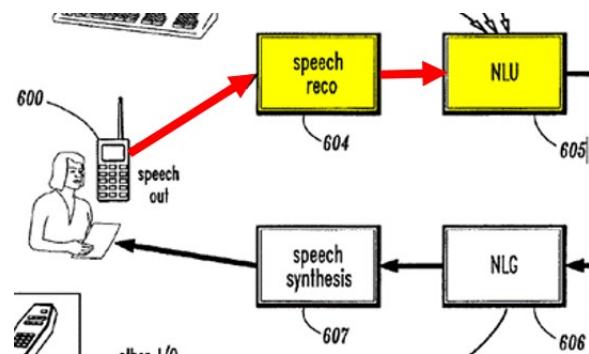
FIG. 6

(Coffman, Fig. 6 (annotations added); *see also id.*, 37:14-22 (“Spoken input from a speech client (e.g., telephone 600). . .”).)

79. Coffman further discloses receiving spoken input that is a “**natural language speech utterance.**” In the context of the ’720 patent, and speech recognition and processing generally, “**natural language**” generally speaking refers to the forms of language used by a person for communication, as opposed to artificial or formal languages such as logics, algebras, and programming languages. (See, e.g., Dafydd Gibbon et al., *Handbook of Multimodal and Spoken Dialogue Systems : Resources, Terminology, and Product Evaluation* (2000) (“Gibbon”) [EX1008],

p.444 (definition of “natural language”).) For example, English is a natural language, while C++ and Java are programming languages. Consistent with this understanding, Coffman discloses several examples of natural language speech utterances, such as “I would like to open this,” “it should be synchronized with my PDA,” and “e-mail from my mother.” (Coffman, 6:32-7:3 (“For example, a mouse click or pointer/stylus pointing action followed by the command ‘I would like to open this’ is disambiguated into a set of attribute value pairs: Command: open, Object: Windows or task selected by the last mouse click.”), 48:1-3 (“Since user identity can be associated in the context of each discourse, command ambiguity can be immediately and transparently resolved (e-mail from my mother is correctly understood, independently of the user).”), 56:25-57:2 (“grocery list,” “it should be synchronized with my PDA,” and “Do it!”).)

80. Coffman’s receipt (and use) of spoken “**natural language**” input is also confirmed by Coffman explaining that the spoken input, once recognized, is subjected to natural language parsing and understanding (“NLU”) processes, as illustrated by Figure 6 (excerpted and annotated at right). (Coffman, e.g., Fig. 6, 53:22-26 (“User inputs can be represented



by their transcriptions or their mappings into a symbolic language after parsing and NLU.”), 53:8-10 (“Voice input from a microphone (e.g. the standard voice input) can be arguments of function calls with the sequence of words, phones, or user identity or queries (symbolic language representation provided by NLU.”), 37:14-22 (“Spoken input from a speech client (e.g., telephone 600) is subject to a speech recognition process 604 and other input (e.g., keyboard, mouse clicks etc) are subject to NLU processing 605.”), 6:6-8 (“In another aspect, the CVM provides the capability to have natural dialog with NLU, NLG, contexts and mixed-initiatives sorted across multiple tasks, processes and discourses (with multiple domains).”), 5:1-4 (“In one aspect of the present invention, the CVM utilizes conversational subsystems (which may be local or distributed) including speech recognition, speaker recognition, text-to-speech, natural language understanding and natural dialog generation engines to understand and generate dialog between and user and machine.”), 37:29-32 (“The key conversational aspect is in the input procedure wherein commands can be entered in NLU (to provide natural language understanding of input queries) or in FSG mode (for constrained input according to rules: grammar and vocabulary, as opposed to free natural input).”).)

81. A person of ordinary skill in the art would further have understood and found it obvious for the “**speech unit**” in Coffman’s conversational computing

system to **“convert[] the received natural language speech utterance into an electronic signal.”** In a speech-processing system using conventional digital computing components, such as Coffman’s conversational system, a person of ordinary skill in the art would have understood that, in order to process a user’s utterance, the utterance would need to be converted into an electronic speech signal. (Coffman, 13:1-3 (“The CVM system may be built on top of a conventional OS and APIs 15 and conventional device hardware 16 and located on a server or any client device (PC, PDA, Pvc).”); EX1009, pp.27 (defining “analog-to-digital converter” as “[a] device that converts a continuously varying (analog) signal, such as sound or voltage, from a monitoring instrument into binary code for use by a computer.”), 159 (defining “digital signal” as “[a] signal, such as one transmitted within or between computers, in which information is represented by discrete states—for example, high and low voltages—rather than by fluctuating levels in a continuous stream, as in an analog signal.”).) Coffman explains, for example, that “[t]he speech is captured (and transformed into features) on the respective networked devices using acoustic signal processing engines (audio subsystems) 232, 232a and suitable audio hardware 233, 233a.” (Coffman, 26:11-13.) As part of the functionality for processing incoming speech, a person of ordinary skill in the art would have understood and found it obvious that the **“speech unit”** would **“convert[] the**

received natural language speech utterance into an electronic signal.”

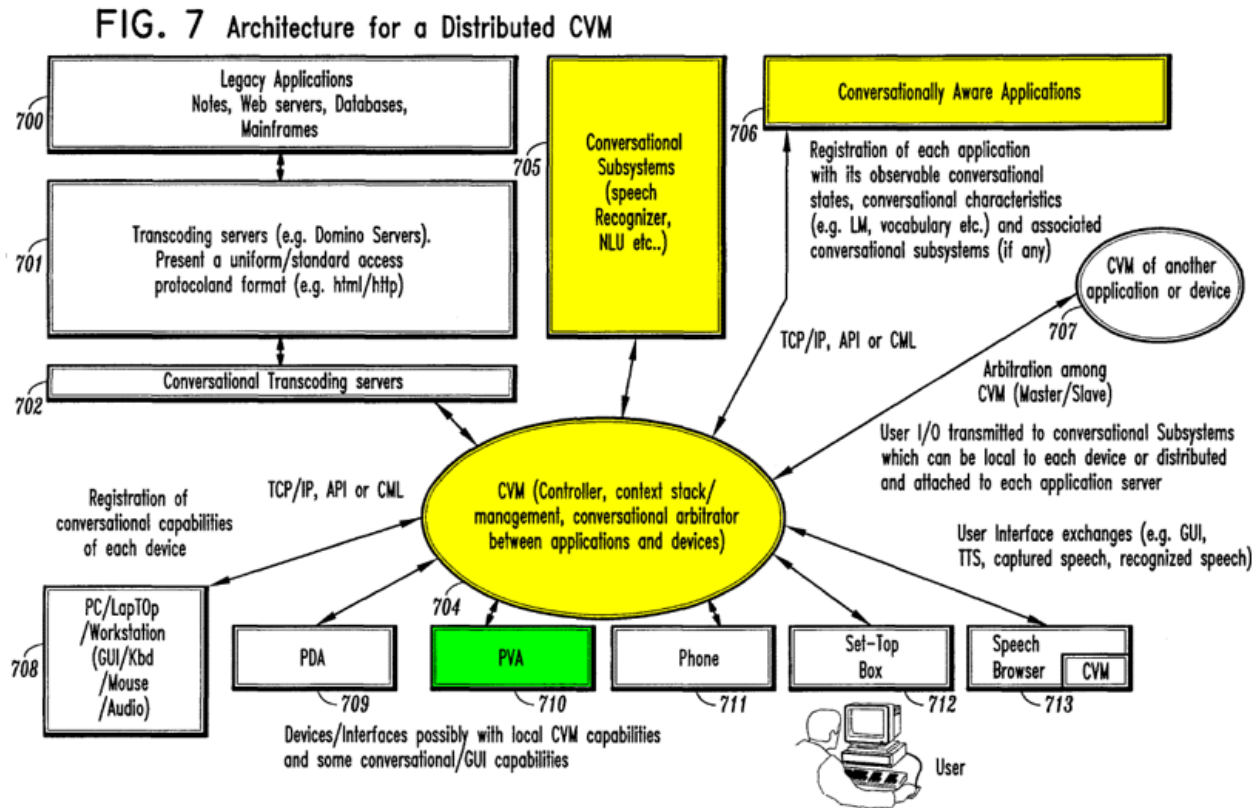
(b) “a natural language speech processing system connected to the computer device on the vehicle, wherein the natural language speech processing system receives, processes, and responds to the electronic signal using data received from a plurality of domain agents, wherein the natural language speech processing system includes:” (1[b])

82. As discussed for 1[a], Coffman discloses **“the computer device on the vehicle”** in the form of a “conversational car PC (PVA, Personal Vehicle Assistant).” (Coffman, 56:21-25 (“[Y]ou have, in your car, a conversational PDA, a conversational car PC (PVA, Personal Vehicle Assistant) and a conversational smart phone. Further assume that the PVA runs an embedded CVM while the two other applications are conversationally aware, i.e., you can control them through the CVM running on the PVA.”), 55:27-28 (“When a user interacts with the collection of devices, the interaction may always be via a central unit such as a PVA (personal vehicle assistant) 710 in a car, or a speech browser 713.”).)

83. In Coffman, the claimed **“natural language speech processing system”** connected to the conversational car PC (PVA) (**“the computer device on the vehicle”**) comprises the conversational system’s CVM, conversational subsystems, and conversational APIs (application program interfaces) and protocols that enable the CVM, subsystems and conversationally-aware applications to

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

communicate with each other. Figure 7 of Coffman illustrates these components and their connection to the PVA:



(Coffman, Fig. 7 (highlighting added).)

84. With respect to Coffman's "conversational subsystems," Figure 4 of Coffman further illustrates how it comprises a speech recognition engine, natural language understanding and natural language parsing system, and a text-to-speech synthesis (TTS) system:

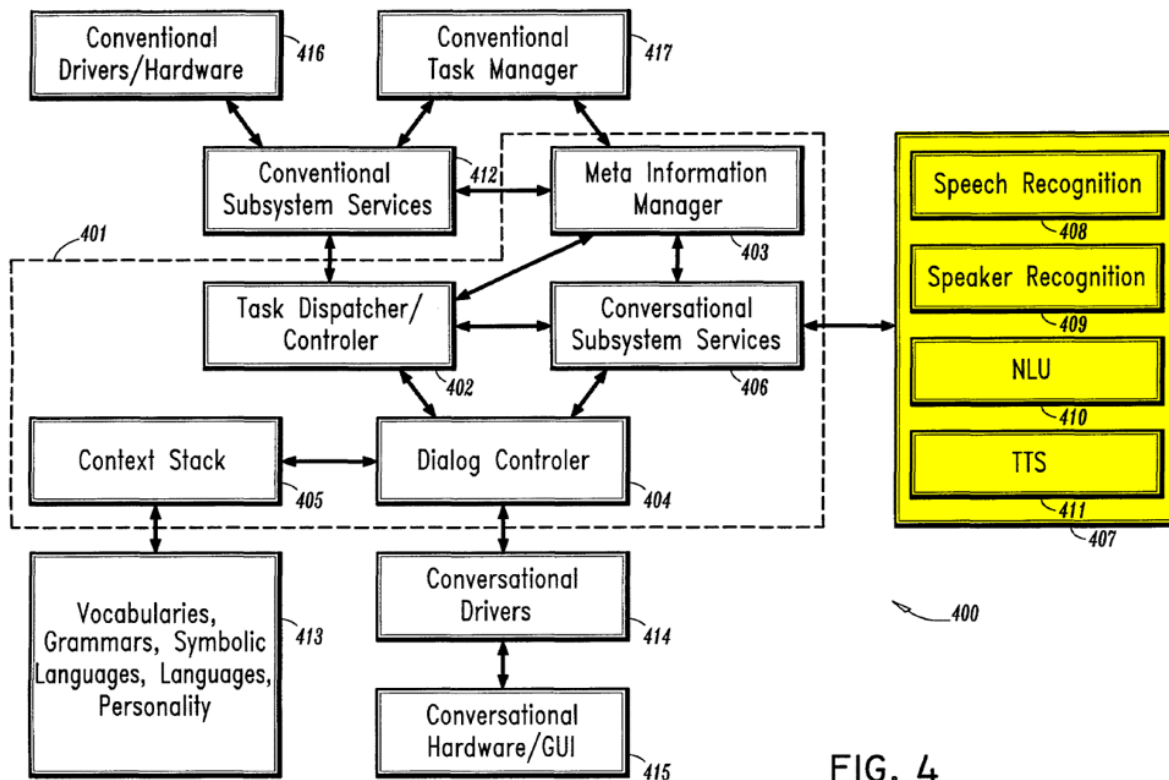


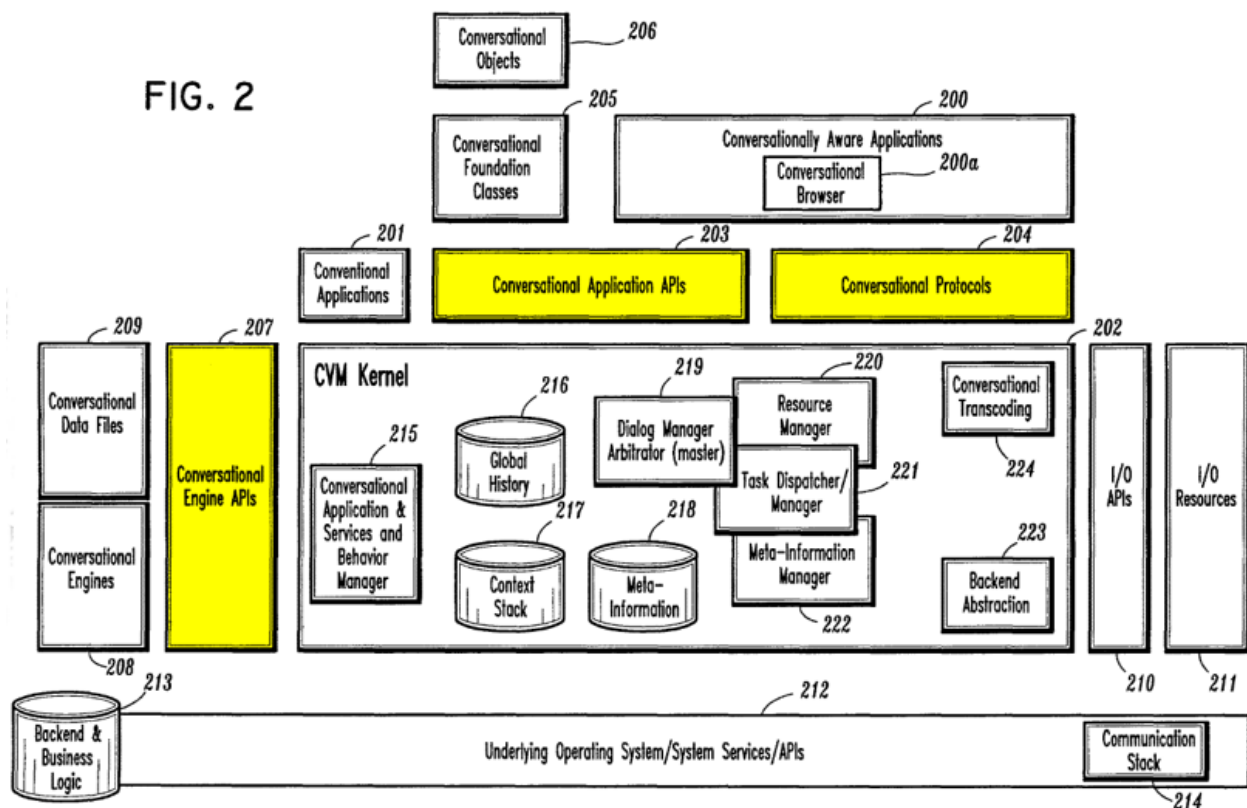
FIG. 4

(Coffman, Fig. 4 (highlighting added); *see also id.*, 29:1-4 (“The conversational system 400 further comprises a plurality of conversational resource subsystems (engines) 407 including, for example, a speech recognition system 408, a speaker recognition system 409, a natural language understanding and natural language parsing system 410 and a text-to-speech (TTS) system 411.”).) I will address Coffman’s speech recognition engine and natural language understanding and natural language parsing system in my analysis of 1[b][i] and 1[b][ii], which recite “a speech recognition engine” and “a parser,” respectively.

85. With respect to the conversational system’s conversational APIs and

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

protocols, Coffman explains that “[i]n general, the CVM provides conversational APIs and protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.” (Coffman, 10:28-30.) Figure 2 of Coffman illustrates:



(Coffman, Fig. 2 (highlighting added).) As shown in the figure above, the APIs include both “conversational application APIs,” which enable a conversationally-aware application to communicate with a CVM, as well as “conversational engine APIs,” which enable communication with conversational subsystems. (Coffman, 5:1-6 (“In one aspect of the present invention, the CVM utilizes conversational

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

subsystems (which may be local or distributed) including speech recognition, speaker recognition, text-to-speech, natural language understanding and natural dialog generation engines to understand and generate dialog between and user and machine. These subsystem are accessed through the CVM. The engines are hidden to the application through the conversational application APIs. The CVM may control such engines through the conversational engine APIs.”), 19:16-19 (“The conversational computing system of Fig. 2 further comprises a conversational engine API layer 207 which provides an interface between core engines conversational engines 208 (e.g., speech recognition, NL parsing, NLU, TTS and speech compression/decompression engines) and the applications using them.”), 12:28-30 (“The conversationally aware applications 11 are applications that are specifically programmed for operating with a CVM core layer (or kernel) 14 via conversational application APIs 13.”), 13:17-28 & 13:29-32 (describing conversational application API layer).) In addition to the APIs, “[t]he conversational protocols 204...comprise conversational discovery (detection), registration, and negotiation protocols (or methods) 301.” (Coffman, 24:10-11.) For example, “[t]he registration protocols allow each networked device or application to exchange and register information regarding their conversational capabilities, state/context and arguments.” (Coffman, 24:11-13; *see also id.*, 22:10-23:3 (“The distributed conversational protocols 300

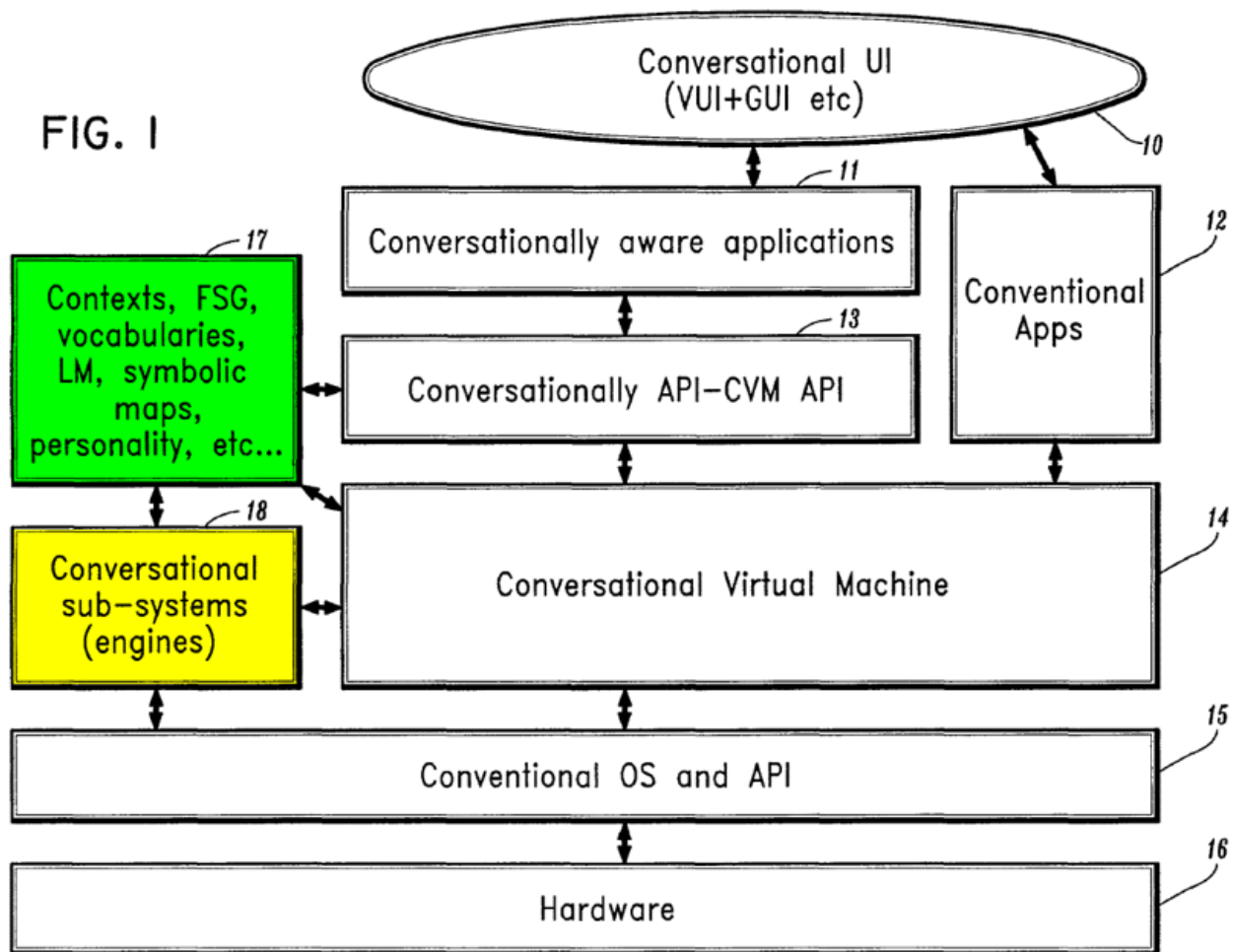
Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

allow network conversational applications 200, 200a and network-connected devices (local client and other networked devices such as a server) to exchange information register their current conversational state, arguments (data files 209) and context with each other.”.)

86. Coffman also makes clear that, as software components running on a conventional computing device, the CVM, conversational subsystems, and conversational APIs and protocols (“**natural language speech processing system**”) would be “**connected to**” the conversational car PC (PVA) (“**the computer device on the vehicle**”). (Coffman, 13:1-3 (“The CVM system may be built on top of a conventional OS and APIs 15 and conventional device hardware 16 and located on a server or any client device (PC, PDA, PvC).”), 29:1-4.) For example, as noted above, Figure 7 of Coffman depicts PVA 710 coupled to these components. Coffman also describes a “PVA run[ning] an embedded CVM” and how a user’s interaction with the conversational system “may always be via a central unit such as a PVA (personal vehicle assistant) 710 in a car.” (Coffman, 56:21-25, 55:27-28.) A person of ordinary skill in the art would have understood and found it obvious that, in order for the PVA to provide conversational computing functionality, the PVA would be connected to these components.

87. Turning to the second portion of 1[b], Coffman’s “**natural language**

speech processing system” (CVM, conversational subsystems, and conversational APIs and protocols) **“receives, processes, and responds to the electronic signal using data received from a plurality of domain agents.”** The claimed **“data received from a plurality of domain agents”** corresponds to data provided by conversationally-aware applications. As noted above, when a conversationally-aware application registers with a CVM system, it provides various data to the system, including its active vocabulary, grammar, and language models for controlling the application. (Coffman, e.g., 57:22-27, 7:12-14, 13:8-11, 24:31-25:9, 36:31-32 (“each application will provide the CVM 401 its own parsing and translation arguments”).) I discuss in more detail in my analysis of 1 [b][i]-[iii] how Coffman’s system **“receives, processes, and responds to the electronic signal using [this data],”** including by performing speech recognition and natural language parsing on the user’s input, selecting the appropriate conversationally-aware application to process the user’s input, and generating output to the user. For example, referring to Figure 1, Coffman explains that “[i]n general, conversational sub-systems 18 are responsible for converting voice requests into queries and converting outputs and results into spoken messages using the appropriate data files 17 (e.g., contexts, finite state grammars, vocabularies, language models, symbolic query maps etc.)” (Coffman, 13:8-11):



(Coffman, Fig. 1 (highlighting added).)

88. Turning to why a conversationally-aware application is a “**domain agent**,” the ’720 patent provides the following description of “agents”:

Agents are executables that receive, process and respond to user questions, queries and commands. The agents provide convenient and re-distributable packages or modules of functionality, typically for a specific domain. Agents can be packages of executable code, scripts, links to information, data, and other data forms, required to provide a

specific/package of functionality, usually in a specific domain. In other words, an agent may include everything that is needed to extend the functionality of the invention to a new domain. Further, agents and their associated data can be updated remotely over a network as new behavior is added or new information becomes available. Agents can use system resources and the services of other, typically more specialized, agents. Agents can be distributed and redistributed in a number of ways including on removable storage media, transfer over networks or attached to emails and other messages. An update manger is used to add new agents to the system or update existing agents.

(’720, 4:7-24.) Generally speaking, in the software context, an “agent” broadly refers to a program that performs actions on behalf of another entity, such as a user or another software program or computing system. For example, a web browser is sometimes referred to more technically as an “HTTP user agent” (HTTP being short for Hypertext Transfer Protocol, a communication protocol typically associated with the World Wide Web). Similarly, desktop email software (such as Microsoft Outlook) is sometimes referred to as a “mail user agent” or MUA.

89. As noted above, a “**domain agent**” in Coffman corresponds to a conversationally-aware application—the executable software and associated data, such as vocabulary, grammar, language models and NLU symbolic mapping. (Coffman, e.g., 13:26-29 (“The code implementing the applications, API calls and

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

protocol calls includes interpreted and compiled scripts and programs, with library links, conversational logic (as described below) engine call and conversational foundation classes.”.) A conversationally-aware application, consistent with the ’720 patent’s description of “agent,” receives, processes, and responds to user questions, queries and commands. (Coffman, e.g., 13:8-13 (queries transformed into “application calls” and application output returned to user), 37:2-4 (task executes query), 40:9-19 (“execute completed queries” and “output of executed or running processes”).) A conversationally-aware application, consistent with the ’720 patent’s description of “agent,” also usually provides functionality related to a specific domain, such as a particular subject matter (such as email or spreadsheet) or device (such as a PDA or smart phone). (Coffman, 16:35-17:7 (listing “application specific” foundation classes, including “Manage_e-mail,” “Manage_calendar,” and “Manage_addressbook/director”), 56:11-14 (e-mail), 58:30-32 (spreadsheet), 56:22-25 (“conversational PDA,” “conversational car PC (PVA, Personal Vehicle Assistant),” and “conversational smart phone”).) Also consistent with the ’720 patent’s description of “agent,” Coffman describes modular development tools for conveniently building complete and re-distributable applications, for example:

Using developmental tools, the developer can easily build his

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

application around the conversational API and CVM. This development environment (referred to herein as Spoken Age) allows programmers to emulate CVM, to debug applications or networked protocols and to rapidly develop conversational user interfaces. Spoken Age includes the CUI and application development for CVM. It also provides the environment for modifying the data files (conversational arguments) of the engines for a given application. In particular this means that at the level of the tools, Spoken Age also includes conventional engine front-ends like SDK Toolkit like the IBM ViaVoice toolkits. This means that toolkits and the algorithms that it provides allows the user to re-build, adapt or extend the data files for a given task.

(Coffman, 58:12-20; *see also id.*, 58:27-28 (“Once an application is developed on a platform and for a specific channel, programmers can rely on Spoken Age to port it to any other platform supporting CVM.”), 7:29-8:9 (“[D]evelopment tools are provided for developer to build, simulate and debug conversational aware application for CVM. The development tools offer direct implementation of the API calls, protocol calls, application using these API’s and protocols, and linking associated libraries, applications exploiting the services and behaviors offered by CVM.... The code implementing these applications can be declarative or procedural. This comprises interpreted and compiled scripts and programs, with library links, conversational logic, engine calls, and conversational foundation classes.

Conversational foundation classes are the elementary components or conversational gestures that characterize any dialog, independently of the modality or combination of modalities.”.) Coffman explains that these applications can be distributed for installation by conversational systems. (Coffman, 52:9-11 & 52:18-21 (describing agents to upgrade or install applications).)

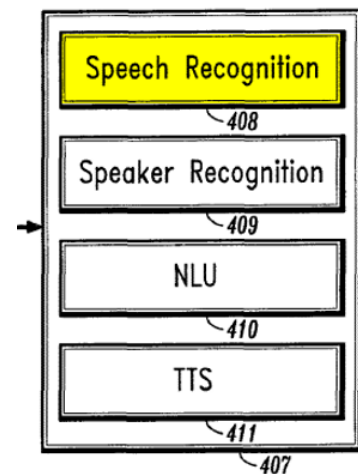
90. A conversationally-aware application is a “**domain agent**” because, as explained, it provides functionality for a particular area or realm, such as a particular subject matter (such as email or spreadsheet) or a device (such as a PDA or smart phone). (Coffman, 16:35-17:7 (listing “application specific” foundation classes, including “Manage_e-mail,” “Manage_calendar,” and “Manage_addressbook/director”), 56:11-14 (e-mail), 58:30-32 (spreadsheet), 56:22-25 (“conversational PDA,” “conversational car PC (PVA, Personal Vehicle Assistant),” and “conversational smart phone”), 36:29-31 (“More specifically, the present invention provides the capability to have a natural dialog with NLU, NLG and mixed initiative across multiple applications, with multiple domains.”).)

91. Lastly, as explained below for 1[b][i]-[iii], it would have been obvious for Coffman’s “**natural language speech processing system**” to include “**a speech recognition engine**” (1[b][i]), “**a parser**” (1[b][ii]) and “**an agent architecture**” (1[b][iii]) as claimed.

- (i) **“a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;” (1[b][i])**

92. As explained below, 1[b][i] would have been obvious over Coffman in view of Kanevsky [EX1004] and Kneser [EX1005].

93. To start, Coffman discloses “a speech recognition engine.” As explained for 1[b], Coffman explains that its system includes various “conversational subsystems,” including a “speech recognition engine” (illustrated at right (Fig. 4 (excerpt))). (Coffman, 10:28-30 (“In general, the CVM provides conversational APIs and protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.”), 29:1-4 (“The conversational system 400 further comprises a plurality of conversational resource subsystems (engines) 407 including, for example, a speech recognition system 408”), Fig. 4.)



94. Coffman further makes clear that its speech recognition engine

“recognizes at least one of words or phrases from the electronic signal,” explaining that “[s]poken input from a speech client (e.g., telephone 600) is subject to a speech recognition process 604 and other input (e.g., keyboard, mouse clicks etc) are subject to NLU processing 605.” (Coffman, 37:20-22; *see also id.*, Fig. 6 (depicting input to and output from “speech reco” 604), 38:9-11 (“Each input stream is converted into its ASCII transcription and aligned with input time- marks via speech recognition processing 604.”).)

95. A person of ordinary skill in the art would have also understood and found it obvious that the speech recognition in Coffman’s system would be performed **“using at least the data received from the plurality of domain agents,”** including **“a plurality of dictionary and phrase entries.”** As explained below, this data in Coffman comprises n-gram language model data used for speech recognition that is provided by conversationally-aware applications.

96. In particular, as noted, Coffman explains that each conversationally-aware application registers with a conversational virtual machine (“CVM”), and provides its language model(s) for controlling the application. (Coffman, e.g., 57:22-58:11, 7:12-14, 13:8-13.) By way of background, in a speech recognition context, a “language model” generally speaking refers to a formal representation of the structure of a natural language. (*See, e.g.*, Gibbon [EX1008], p.430 (definition

of “language model”).) “In most speech recognisers language models consist of an n-gram model, i.e. probabilities of the occurrence of words, words pairs, sequences of three words, etc.” (Gibbon, p.430; *see also id.*, pp.444 (defining “n-gram language model” as “a stochastic language model that is based on a n-gram grammar, i.e. on conditional probabilities depending only on the (n-1) immediate predecessor words.”), 479 (defining “stochastic language model” as “a language model that assigns probabilities to the allowed word sequences; typically all word sequences have a non-zero probability.”).) A “trigram” language model therefore refers to an n-gram language model that uses the probabilities of the occurrence of sequences of three words (i.e., the “n” in “n-gram” is three). (Gibbon, p.488 (defining “trigram” as “a sequence of three words”).)

97. Coffman does not provide a detailed discussion of how a language model is used for speech recognition. This is not surprising, because speech recognition using language model(s) was well known and I understand that a patent need not teach, and preferably omits, what is well known in the art. Coffman does expressly incorporate by reference U.S. Patent No. 5,897,616 (Kanevsky [EX1004]) (Coffman, 37:10-13), which further explains that a speech recognition engine may be implemented using “a trigram language model.” (Kanevsky, 13:50-56.) Kanevsky describes such a speech recognition engine as an “existing component”

and explains that “[s]uch systems, for example, are disclosed in IBM’s speech dictation engine ‘Simply Speaking’ and in ‘Transcription of Radio Broadcast News with the IBM Large Vocabulary Speech Recognition System’, P.S. Gopalakrishnan [sic, Gopalakrishnan] et al., Proceeding Speech Recognition Workshop, Arden House, Drapa (1996).” (Kanevsky, 13:48-61.)

98. Under the proposed combination of Coffman and Kanevsky, therefore, the speech recognition process to identify one or more trigrams corresponding to the user’s utterance discloses **“recogniz[ing] at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries”** because the trigrams comprise words and sequences of words (i.e., phrases). (*See* Gibbon [EX1008], p.488 (defining “trigram” as “a sequence of three words”).) The trigram data in a language model comprises **“a plurality of dictionary and phrase entries”** at least because a person of ordinary skill in the art would have understood and found it obvious to organize the n-gram language model(s) used by the Coffman’s conversational system as a table, with the trigram data being **“entries”** in the table. For example, Sutcliffe [EX1016] explains that “[m]ost large-vocabulary English speech recognition systems use a word N-gram language model of English grammar: syntactic

knowledge is captured in a large table of word bigrams (pairs), trigrams (triples), ... N-grams (see surveys of large-vocabulary speech recognition systems, e.g. HLT Survey, 1995; comp.speech, 1996).” (Richard F. E. Sutcliffe et al., *Industrial Parsing of Software Manuals* (1996) (“**Sutcliffe**”) [EX1016], p.27.) U.S. Patent Application Publication No. 2003/0144846 (“Denenberg”) (on which I am named as a co-inventor) similarly explains that “[a]nother common recognizer is a natural language speech recognizer based upon the N-gram language model which works from tables of probabilities of sequences of words.” (Denenberg [EX1014], ¶0040.) To form an n-gram, the table contains an entry for a particular word, with a link to other word(s). Thus, a trigram language model comprises “**a plurality of dictionary and phrase entries**” because it is organized as a table and includes a list of words and sequences of words to refer to during speech recognition. (See, e.g., Merriam-Webster’s Collegiate Dictionary (10th ed. 1994) [EX1015], p.322 (defining “dictionary” as encompassing “a list (as of items of data or words) stored in a computer for reference (as for information retrieval or word processing)”).)

99. **Rationale and Motivation to Combine (Coffman and Kanevsky).** It would have been obvious to a person of ordinary skill in the art to combine Coffman and Kanevsky, predictably resulting in the conversational system of Coffman in which language models provided by the conversationally-aware applications and

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

used by the system for speech recognition were n-gram language models, such as a trigram language model. Combining the teachings of Coffman and Kanevsky in this manner would have therefore resulted in **“recogniz[ing] at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries”** as claimed.

100. Both Coffman and Kanevsky are analogous references in the same field of speech recognition and processing as the '720 patent. ('720, 1:12-21; Coffman, e.g., 1:7-8 (“systems and methods for conversational computing”), Figs. 1, 4 & 6; Kanevsky, e.g., 1:15-21 (“automatic speech recognition, text-independent speaker identification, natural language understanding techniques”).) Coffman and Kanevsky would have also been reasonably pertinent to a number of problems facing the patent inventors, including how to implement a system for a person to interact with a computer system using speech and techniques for performing speech recognition.

101. A person of ordinary skill in the art would have been motivated to combine Coffman and Kanevsky. To begin with, as noted above, Coffman expressly incorporates Kanevsky by reference. (Coffman, 37:10-13.) Therefore, I am informed that Kanevsky is properly considered part of the disclosures of Coffman

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

for purposes of applying it as prior art. I further note that both references also share a named co-inventor, Stephane Herman Maes, and identify IBM as the assignee/applicant. (Coffman, face page; Kanevsky, face page.) Therefore, a person of ordinary skill in the art would have naturally consulted both of these references to understand the features and functionality of Coffman's conversational system, or to actually implement or adapt the system and features it provides. A person of ordinary skill in the art would have regarded Coffman and Kanevsky as a collective body of work, with Kanevsky complementing Coffman's disclosures. Because Coffman incorporates Kanevsky by reference, and both references share a common co-inventor and assignee, and in light of their overlapping subject matter, a person of ordinary skill in the art would have found it odd to even question their combinability, or whether they can properly be considered together to understand Coffman's conversational system.

102. Moreover, as explained, Coffman already discloses that a conversationally-aware application provides its "language models" to the conversational system in connection with processing a user's input. A person of ordinary skill in the art would have understood that a language model in this context likely refers to an n-gram language model (*e.g.*, Gibbon [EX1008], pp.430 (definition of "language model"), 444 (definition of "n-gram language model")), and

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

Kanevsky provides express confirmation, explaining that a speech recognition engine can be implemented using a trigram language model (Kanevsky, 13:50-56). Moreover, n-gram language models were one of a number of finite ways of implementing speech recognition, and the use of an n-gram language model as taught by Kanevsky would have involved the application of a known technique to implement or improve Coffman's speech recognition functionality in the same way.

103. A person of ordinary skill in the art would have had a reasonable expectation of success. The implementation and use of n-gram language models for speech recognition was well-known and would have involved the application of conventional techniques and technologies. For example, as noted, the '720 patent itself acknowledges that "[s]peech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications." ('720, 1:58-59.) Kanevsky describes a speech recognition engine using a trigram language model as an "existing component," disclosed in, for example, "IBM's speech dictation engine 'Simply Speaking' and in 'Transcription of Radio Broadcast News with the IBM Large Vocabulary Speech Recognition System', P.S. Gopalakrishnan [sic, Gopalakrishnan] et al., Proceeding Speech Recognition Workshop, Arden House, Drapa (1996)." (Kanevsky, 13:48-61.) A person of ordinary skill in the art would have found the disclosures of Coffman and Kanevsky sufficiently detailed to

implement the techniques on which I rely, without undue experimentation.

104. Claim 1 [b][i] further recites that the plurality of dictionary and phrase entries “**are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user.**” As explained next, this portion of 1 [b][i], describes nothing more than the conventional use of an “adaptive” n-gram language model (such as bigram or trigram) for speech recognition. In an adaptive n-gram language model, the probabilities associated with the n-grams are dynamically updated as the speech recognition system is used. (*See, e.g.*, Gibbon [EX1008], p.371 (defining “adaptive language model” as a language model that “adapts [its] probabilities to the most recent history, say the last 100 to 1000 predecessor words.”).)

105. Coffman itself suggests modifying language models based on user interaction, describing collection of user-specific information in order to perform “more robust recognition,” and explaining that “any information can be accumulated to recognize the user,” including “the usual phrasing of a query, the type of query phrased, command frequency (often used, not often used), preferred applications, time or usage, etc.” (Coffman, 37:5-10.) Although neither Coffman nor Kanevsky appears to expressly detail using them, adaptive n-gram language models and their advantages were well known. The use of adaptive n-gram language models would

have been obvious over U.S. Patent No. 6,157,912 (**Kneser** [EX1005]). Kneser discloses that it is advantageous to update a language model's probabilities that "given word sequences of limited length" (i.e., n-grams) are present based on the user's interaction with the system, explaining:

These language model values represent the probabilities that given word sequences of limited length are present; they have been obtained from a comparatively large quantity of training texts, which may also be formed by written texts. Long-term changes in the use of language, for example subject matter or style, can accordingly not be taken into account with these language model values [probabilities]. Since the quality of the adaptation of the language model values has a considerable influence on the reliability of the speech recognition, it is desirable to adapt the language model values as well as possible to the given text by means of a small quantity of updated text. This adaptation may take place during the process of recognizing a given text or alternatively subsequent to the recognition of part of a text, after recognition errors have been manually corrected.

(Kneser, 1:22-29; *see generally id.*, 1:13-37.)

106. **Rationale and Motivation to Combine (Coffman and Kanevsky with Kneser)**. It would have been obvious to a person of ordinary skill in the art to further combine Coffman and Kanevsky with Kneser's teachings regarding updating language model n-gram probabilities as the user interacts (e.g., dialogs) with the

system. The accumulation of language model updates as the user dialogs with the system over time would have predictably resulted in language models in Coffman's conversational system having dictionary and phrase entries that **“are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user.”**

107. Like Coffman and Kanevsky, Kneser is an analogous reference in the same field of speech recognition and processing as the '720 patent. ('720, 1:12-21; Kneser, e.g., 1:5-11 (“The invention relates to the automatic recognition of speech in which a language model with language model values are used, which values are determined from a comparatively long training text, while an adaptation of the language model is carried out for certain cases, in particular for certain types of texts.”).) Kneser would have also been reasonably pertinent to a number of problems facing the patent inventors, including techniques for performing speech recognition and how to implement a speech recognition system that uses “dynamic” probability information.

108. A person of ordinary skill in the art would have been motivated to combine Kneser with Coffman and Kanevsky. Kneser, as shown above, provides express motivation for incorporating “adaptive” language model techniques into Coffman's conversational system—improving speech recognition reliability and

performance. Coffman itself confirms this motivation in the context of collecting user-specific information in order to perform “more robust recognition”:

It is to be appreciated that conversational biometrics can be used to collect any context and meta information on the user not only to customize or adapt for purposes of user preferences or to authorize a query, but also to use the information to perform more robust recognition. Accordingly, any information can be accumulated to recognize the user. Namely, the usual phrasing of a query, the type of query phrased, command frequency (often used, not often used), preferred applications, time or usage, etc.

(Coffman, 37:5-10.) Adaptation based on usual phrasing and frequency of commands, as Coffman describes above, could be easily accomplished by adjusting the probabilities in the n-gram language model, and adaptive language models as taught by Kneser provide a way in which Coffman’s system could advantageously improve speech recognition reliability and performance. The use of adaptive language models was well known in the art (*e.g.*, Gibbon [EX1008], p.371 (defining “adaptive language model”)), and a person of ordinary skill in the art would have appreciated that using adaptive language models in Coffman’s conversational system for speech recognition would have improved a similar system in the same way.

109. A person of ordinary skill in the art would have had a reasonable

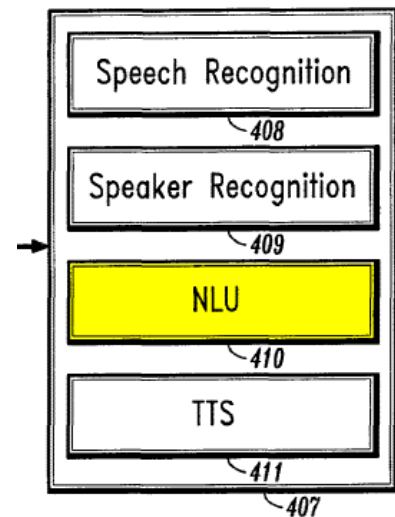
expectation of success. As noted, the use of adaptive language models was well known in the art, and implementing the language models used in Coffman's conversational system as adaptive language models would have involved conventional techniques and technologies. A person of ordinary skill in the art would have found the disclosures of Coffman, Kanevsky, and Kneser sufficiently detailed to implement the techniques on which I rely, without undue experimentation.

- (ii) **“a parser that interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases, wherein the parser interprets the recognized words or phrases by:” (1[b][ii])**

110. As explained below, 1[b][ii] and its sub-steps (1[b][ii][1]-[3]) would have been obvious in further view of Zadrozny [EX1006].

111. Coffman discloses “a parser.” As discussed for 1[b], Coffman’s conversational system includes various “conversational subsystems,” including “a natural language understanding and natural language parsing system [NLU] 410”

(illustrated at right (Fig. 4 (excerpt))). (Coffman, 29:1-4 (“The conversational system 400 further comprises a plurality of conversational resource subsystems (engines) 407 including, for example, a speech recognition system 408, a speaker recognition system 409, a natural language understanding and natural



language parsing system 410 and a text-to-speech synthesis (TTS) system 411.”), Fig. 4.) Coffman explains that, “using NLU technology,” “[q]ueries can be decoded, parsed and then translated into a logic combination of queries (symbolic query).” (Coffman, 51:12-14; *see also id.*, 40:12-15 (“By way of example, the task dispatcher/controller 402 will decide on the basis of the context stack 405 the different statistical parsers that must operate on a query for the dialog controller 404 to identify the active context and complete the query.”).)

112. The remainder of 1[b][ii] introduces and summarizes sub-steps 1[b][ii][1]-[3]. Therefore, I explain below in my analysis of 1[b][ii][1]-[3] how Coffman’s natural language understanding and natural language parsing system

(“parser”) “interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases.” In short, as I explain, Coffman teaches that its NLU processing uses data provided by conversationally-aware applications (such as grammars and symbolic query maps) to interpret the words or phrases in the user’s utterance in order to identify an appropriate application and to formulate an application call/query to perform. (Coffman, e.g., 6:32-7:3, 13:8-11, 30:8-10, 37:20-26, 41:2-5, 42:1-5, 51:12-14, 53:25-26.)

(1) “determining a context for the natural language speech utterance;” (1[b][ii][1])

113. A person of ordinary skill in the art would have understood and found it obvious that Coffman’s parser “interprets the recognized words or phrases by” **“determining a context for the natural language speech utterance.”** As explained below, Coffman uses NLU (i.e., parsing) as part of the process to determine a “context” that matches the user’s recognized utterance.

114. Coffman explains that the conversational system includes a “context stack” where the state and history of each active process, such as an application process, is tracked in its own “context”:

The core of the CVM 401 is the context stack 405 which operates and is managed under the control of the dialog controller 404 (it is to be

understood that the context stack 405 id [sic; is] directly related to the global history and meta information repository discussed above). In general, the context stack 405 accumulates the context (i.e., full query arguments list of attribute value n-uples, and state/mode) of each active process with an activated discourse (i.e., conversational interaction associated with a given task/process/thread) along with any data files 413 (or at least identifiers of such conversational arguments) for the different engines that may be needed for input understanding (e.g., files or arguments that the engines use for performing their respective tasks such as active FSG, topic, vocabulary, HMM (hidden markov models), voiceprints, language models or possible queries for a speech input). In other words, the term “context” refers to the state of each discourse (whether active or normative), which keeps track of the past history of the discourse, its current state, and the specific characteristics and full query arguments of the corresponding task (e.g, vocabulary file, language model, parsing, tags, voiceprint, TTS rules, grammar, NLU etc. of each active task/process) along with any data needed for input understanding. It is to be appreciated that, in a distributed CVM (as explained below), the context stack (as well as other CVM components) may be directly associated with networked services (i.e., distributed over the network) (as described above with respect to the context and global history).

(Coffman, 29:18-30:3; *see also id.*, 30:4-10 (“More specifically, each new task, process, or thread creates a new stack entry and is associated with a discourse. Each

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

application may be associated with multiple discourses (e.g. the application management discourse and the application content navigation discourses). Each context associated with a given discourse comprises the latest requests made to the corresponding process/task/thread as well as the latest output. The context of a given discourse is also associated with, e.g., any active grammars, vocabularies and symbolic language which maps the actual query.”.)

115. Coffman further discloses that, in order to determine a “context” for processing a user’s request/command, NLU processing (i.e., parsing) is used to match recognized words in the user’s utterance to a context. Coffman explains, for example:

Completion/search for the active context is performed from context to context down the stack. That is, new queries or arguments are compared by the dialog engine by going down the stack until an acceptable match is obtained and optionally confirmation is obtained from the user. As soon a context is found that fits the utterance at the level of the NLU symbolic language, the context becomes active and the corresponding process becomes active.

(Coffman, 42:1-5; *see also id.*, 41:2-5 (“Based on the modality (pointer, keyboard, file, speech), the task dispatcher 402 redirects the stream to the appropriate conversational subsystems or conventional subsystem with speech inputs being transcribed/understood. The output of these subsystems is run down the context

stack 405 to extract the active query and complete it.”), 30:8-10 (“The context of a given discourse is also associated with, e.g., any active grammars, vocabularies and symbolic language which maps the actual query.”), 51:12-14 (“Queries can be decoded, parsed and then translated into a logic combination of queries (symbolic query) using NLU technology.”), 40:12-15 (“By way of example, the task dispatcher/controller 402 will decide on the basis of the context stack 405 the different statistical parsers that must operate on a query for the dialog controller 404 to identify the active context and complete the query.”), 37:20-26 (“Each input is subject to attribute acquisition (401a) whereby the attribute value n-uples are acquired from the input. A summarization process 401b is performed whereby the attribute value n-uples are added to the context and then verifies with the syntax of the back-end application 608 whether the query is complete, incomplete, or ambiguous.”).)

**(2) “selecting at least one of the plurality of domain agents based on the determined context; and”
(1[b][ii][2])**

116. A conversationally-aware application (“**domain agent**”) is “**select[ed]...based on the determined context**” because activation of an application for processing the user’s query is based on the application’s association with the determined context. Coffman explains:

Completion/search for the active context is performed from context to context down the stack. That is, new queries or arguments are compared by the dialog engine by going down the stack until an acceptable match is obtained and optionally confirmation is obtained from the user. As soon a context is found that fits the utterance at the level of the NLU symbolic language, the context becomes active and the corresponding process becomes active.

(Coffman, 42:1-7; *see also id.*, 37:2-4 (“The first task to have its dialog completed by producing a non-ambiguous query is executed and the corresponding query as interpreted by the other application is stored to activate if the recognized query is rejected by the user.”).)

- (3) **“transforming the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command; and” (1[b][ii][3])**

117. Coffman discloses **“transforming the recognized words or phrases into at least one of a question or a command.”** As discussed, Coffman describes using natural language understanding (parsing) to translate a recognized utterance into a symbolic mapping and query/call to be carried out by an application. (Coffman, e.g., 13:8-13 (“The conversational application API 13 conveys all the information for the CVM 14 to transform queries into application calls and

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

conversely converts output into speech, appropriately sorted before being provided to the user.”), 30:8-10 (“Furthermore, the context of a given discourse is associated with, e.g., any active grammars, vocabularies and symbolic language (attribute value n-uple) which maps the actual query.”), 6:8-10 (“A conversational input interface is provided whereby a set of multi-mode input streams are each transcribed into an ASCII command or query (i.e., lists of attribute-value pairs or n-uples).”).) For example, Coffman teaches that a user’s utterance “I would like to open this” would be processed into “a set of attributes value pairs: Command: Open, Object: Window” and result in an “open” command being performed. (Coffman, 6:32-7:3.)

118. Coffman further discloses and renders obvious **“wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command.”** This because Coffman, as noted, explains that an application (when registering with a conversational system), provides its active vocabulary, grammar, and language models to the CVM for controlling the application. (Coffman, e.g., 24:31-25:9, 57:22-27, 36:31-32 (“each application will provide the CVM 401 its own parsing and translation arguments”).) A person of ordinary skill in the art would have understood and found it obvious that, in a system such as Coffman’s involving speech recognition using language model(s) and natural language understanding, a

grammar associated with an application would be used to formulate the user's question or command as a symbolic mapping (such as a set of attribute-value pairs such as "Command: Open, Object: Window") representing the system's understanding and to, in turn, formulate an application query/call in accordance with that understanding. Both the symbolic mapping and the application query/call would separately correspond to the claimed **"at least one question or command"** **"formulated in a grammar that the selected domain agent uses to process the formulated question or command."**

119. For example, as noted, Coffman describes parsing in connection with natural language processing. (Coffman, 53:25-26 ("User inputs can be represented by their transcriptions or their mappings into a symbolic language after parsing and NLU."), 29:1-4 ("a natural language understanding and natural language parsing system 410").) A person of ordinary skill in the art would have understood and found it obvious in this context that parsing involves using a grammar to identify the parts of speech of the words and phrases in the recognized speech in order to semantically interpret the speech. For example, Gibbon describes "Natural Language Processing" as "[t]he parsing of written language into semantic concepts," where "parsing" is "[t]he labelling of the parts of speech or grammatical elements of sentences, e.g., subject, predicate, past tense, noun, verb, either by a human analyst

or by means of a parser (a programme based on a parsing algorithm).” (Gibbon [EX1008], pp.444, 451.) Coffman’s example of translating the spoken request “I would like to open this” into “a set of attributes value pairs: Command: Open, Object: Window” illustrates parsing of speech in accordance with a grammar. (Coffman, 6:32-7:3.)

120. **Zadrozny** [EX1006] confirms that it would have been obvious to formulate a user’s question or command “**in a grammar that the selected domain agent [conversationally-aware application] uses to process the formulated question or command.**” **Zadrozny**, entitled “NL Understanding with a Grammar of Constructions,” is referenced in Kanevsky as an example of a semantic analyzer. (Kanevsky, 14:22-25.) Zadrozny describes a “grammar of constructions” associated with an application (such as a calendar application) for performing natural language understanding to translate a recognized utterance into “a form that can be executed by the application”:

We present an approach to natural language understanding based on a computable grammar of constructions. A construction consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system. . . .

With respect to parsing, grammars of constructions can be parsed like

“standard” grammars, except that the set of features is richer. Given a string (representing a sentence, a fragment of a discourse or a paragraph), the parser assigns it a construction.... Here the message of an utterance must be understood in the context of an intended action. This is done in two steps. First, the system determines the intended action and its parameters, using domain knowledge (meetings+time+places). Second, once all the parameters have been extracted from the dialog, the system executes the action. To do this, the program uses application-specific knowledge to translate the action and its parameters into a form that can be executed by the application (Xdiary).

(Zadrozny, pp.1-2.) As explained above in the excerpt above, Zadrozny describes a technique using a “grammar of constructions” to formulate an utterance as an intended action and parameters and then translate the action and parameters into a form for the application to execute. Zadrozny illustrates the technique with the example utterance “I want you to arrange a conference in my office at 5.” (Zadrozny, p.3.) Using a grammar of constructions, the utterance is “analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, verb, an object and a complement” and formulated as an action (“schedule”) and series of parameters:

```
[ [ action_name schedule]
  [ event_name
    [ a conference]
  [ event_time
    [ [ minute 0]
```

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

```
        [    hour
          [    5 am_or_pm]
    [    event_place
      [    my office]
```

(Zadrozny, pp.3-4.) As Zadrozny explains, this action and parameters are then translated into a form for the application to execute (i.e., an application call/query).

121. Zadrozny's process is just like Coffman's descriptions of formulating a user's utterance as a set of attribute-value pairs (e.g., "Command: Open, Object: Window") and then converting them into an application call/query. (*See also* Coffman, 30:8-10 ("symbolic language (attribute value n-uple) which maps the actual query"), 13:8-13 (CVM system transforms queries into "application calls").)

122. **Rationale and Motivation to Combine (Coffman, Kanevsky and Kneser with Zadrozny)**. It would have been obvious to a person of ordinary skill in the art to combine Zadrozny's teachings regarding processing speech using a grammar of constructions with the conversational computing system taught by the combination of Coffman, Kanevsky and Kneser. The combination would have predictably resulted in the conversational system in which a user question or command was formulated in a grammar used by a conversationally-aware application to process the formulated question or command.

123. Zadrozny is an analogous reference in the same field of speech recognition and processing as the '720 patent. ('720, 1:12-21; Zadrozny, e.g., p.1

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

(“We present an approach to natural language understanding based on a *computable grammar of constructions*.”) (emphasis in original).) Zadrozny would have also been reasonably pertinent to a number of problems facing the patent inventors, including techniques for understanding natural language speech input and how to determine and formulate actions for an application to undertake in response to the input.

124. A person of ordinary skill in the art would have been motivated to combine. For starters, Kanevsky (which Coffman incorporates by reference) expressly cites Zadrozny as an example of a “semantic analyzer.” (Kanevsky, 14:22-25.) And as noted, authors of Zadrozny are identified as being associated with IBM, the assignee/applicant for both Coffman and Kanevsky. A person of ordinary skill in the art therefore would have naturally turned to Zadrozny for detail about how to implement the systems and techniques described in Coffman and Zadrozny. As with Kanevsky, in fact, a person of ordinary skill in the art would have found it odd to even question Zadrozny’s combinability with Coffman and Kanevsky, or whether Zadrozny can properly be considered to understand Coffman’s conversational system.

125. A person of ordinary skill in the art would have appreciated that the disclosures of Coffman and Zadrozny confirm their natural fit. As noted, Coffman

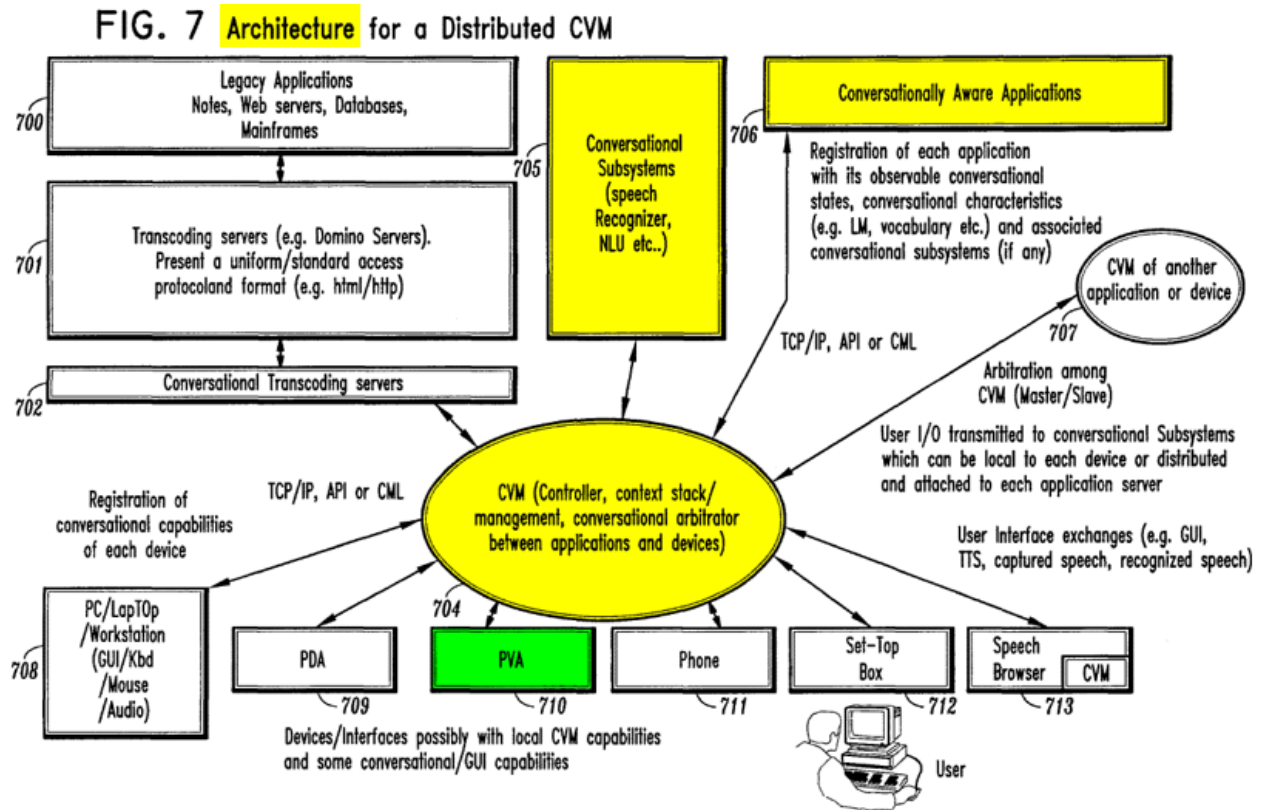
does not describe the natural language understanding process in detail, mostly describing at a high level that applications provide grammars, vocabularies, language models and symbolic mappings, that a context is matched by “fit[ting] the utterance at the level of the NLU symbolic language” and that “CVM 14 [is used] to transform queries into application calls.” (Coffman, 42:4-5, 13:11-13.) Because Kanevsky expressly cites Zadrozny as an example of a semantic analyzer, a person of ordinary skill in the art would have turned to it for its teachings regarding the use of grammars in natural language understanding to translate a recognized utterance into an application call. Moreover, Zadrozny’s process is just like Coffman’s descriptions of formulating a user’s utterance as a set of attribute-value pairs (e.g., “Command: Open, Object: Window”) and then converting them into an application call/query. (*See also* Coffman, 30:8-10 (“symbolic language (attribute value n-uple) which maps the actual query”), 13:8-13 (CVM system transforms queries into “application calls”).)

126. A person of ordinary skill in the art would have had a reasonable expectation of success. Zadrozny, for instance, describes an actual working implementation of a conversational system using a grammar of constructions. (Zadrozny, e.g., pp.3 (describing MINCAL), 5 (“‘complete’ working system”).) A person of ordinary skill in the art would have found the disclosures of Coffman,

Kanevsky, Kneser, and Zadrozny sufficiently detailed to implement the techniques on which I rely, without undue experimentation.

- (iii) **“an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.” (1[b][iii])**

127. Coffman’s **“natural language speech processing system”** (CVM, conversational subsystems, and conversational APIs and protocols) also includes **“an agent architecture”** as claimed. The **“agent architecture”** in Coffman comprises the conversational APIs and protocols that enable communication between the CVM, conversational subsystems, and conversationally-aware applications. Figure 7 of Coffman depicts an example CVM “architecture”:



(Coffman, Fig. 7 (highlighting added).) As discussed for 1[b], Coffman explains that the CVM, conversational subsystems and conversationally-aware applications are communicatively coupled using conversational APIs and protocols. (Coffman, e.g., 10:28-30 (“In general, the CVM provides conversational APIs and protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.”), 5:1-6, 12:28-30, 13:17-28, 13:29-32, 19:16-19, 22:10-23:3, 24:10-11, Fig. 2.)

128. 1[b][iii] recites that the agent architecture “**communicatively couples services of each of an agent manager, a system agent, the plurality of domain**

agents, and an agent library.” As explained above, “**the plurality of domain agents**” corresponds in Coffman to conversationally-aware applications. As discussed below, the “**agent manager**” is disclosed in Coffman by CVM software functionality for a conversationally-aware application to register with a CVM, the “**system agent**” by CVM software functionality including the “CVM kernel,” and the “**agent library**” by conversational subsystems.

129. Starting first with the “**agent manager**,” the written description of the ’720 patent generally describes a “manager” as corresponding to software. (’720, 3:20-28 (“One aspect of the invention includes software that is installed onto the computer, where the software includes one or more of the following modules: ... an event manager for managing events and other modules.”).) The “**agent manager**,” as its name suggests, is described as performing operations for managing agents, such as loading agents into the system. (’720, 21:67-22:6 (“When the system starts-up or boots-up, the agent manager **154** may load and initialize the system agent **150** and the one or more domain agents **156**. At shutdown the agent manager may unload the agents. The agent manager may also perform license management functions for the domain agents and content in the databases **102**.”).)

130. Coffman similarly includes CVM software functionality to manage registration of conversationally-aware applications with the system, i.e., an “**agent**

manager.” Coffman explains that a CVM includes a “conversational application & behavior/service manager layer 215” that “comprises functions for managing the conventional and conversationally aware applications 200 and 201,” including “keeping track of which applications are registered” and “the state of each application.” (Coffman, 19:26-20:2.)

131. Turning next to the “**system agent**,” the written description of the ’720 states that a system agent “provides default functionality and basic services” and “include[s] general-purpose behaviors and data, which provide core or foundation services for more specialized domain or system agents.” (’720, 21:40-41, 4:27-29.) In Coffman, the “**system agent**” is disclosed by CVM software including the “CVM kernel.” The CVM kernel, which Coffman also refers to as the CVM controller, “keeps track of [conversational] resources, transmit[s] input to the appropriate conversational subsystems and arbitrate[s] between devices, state and applications. The CVM kernel layer also coordinates the output generation and prioritization according to the active conversation and conversation history, delayed returns, delegation across network resources and task delegation and memorization.” (Coffman, 5:12-24; *see also id.*, 5:12 (“CVM kernel layer (or CVM controller)”).) This CVM software is a “**system agent**” because, as explained below, it provides default functionality and basic services of the conversational system in order to

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

perform operations on behalf of both a user as well as conversationally-aware applications.

132. For example, Coffman explains that “[t]he core CVM kernel layer 202 comprises programming layers such as a conversational application & behavior/service manager layer 215,” which Coffman describes as providing “core functions” of the CVM kernel layer. (Coffman, 19:26-29.) Coffman explains that the conversational application & behavior/service manager layer 215 comprises functions for managing conversationally aware applications, such as “all the behaviors and features of a conversational UI that the user may expect to find in the applications and interactions” and “the features that an application developer would expect to be able to access via APIs.” (Coffman, 19:26-20:27; *see also id.*, 20:8-14 (listing examples of “conversational services and behavior provided by the CVM kernel” including “output formatting and presentation”).)

133. In addition, the CVM kernel handles input from/output to the user, such as by providing text or speech output from a conversationally aware application to the user. Coffman explains, for example, that I/O resources (such as a touch screen and audio subsystem) will register with the CVM kernel, and that the CVM “automatically handles all input/output issues.” (Coffman, 19:20-25 (“An I/O API layer 210 provides an interface with conventional I/O resources 211 such as a

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

keyboard, mouse, touch screen, keypad, etc. (for providing a multi-modal conversational UI) and an audio subsystem for capturing speech I/O (audio in/audio out). The I/O API layer 210 provides device abstractions, I/O abstractions and UI abstractions. The I/O resources 211 will register with the CVM kernel layer 202 via the I/O API layer 210.”), 13:5-8 (“The CVM automatically handles all the input/output issues, including the conversational subsystems 18 (i.e., conversational engines) and conventional subsystems (e.g., file system and conventional drivers) of the conventional OS 15.”), 58:8-9 (“Again, the CVM automatically handles all the input/output issues, including the conversational and conventional subsystems.”).) Coffman further explains that the CVM’s default input/output functionality can be overridden by, for example, “redirecting” the results provided by an application:

[C]omplete and non-ambiguous commands will result in certain results (e.g., outputs or actions). These results are similarly returned to the appropriate conversational engine 407 with a priority level and update the context, unless if re-directed by the user as in conventional systems. However, the re-direction can be more sophisticated as it can involve partial mixed initiative notification while re-directing the results.

(Coffman, 41:23-26; *see also id.*, 43:6-7 (“Therefore, an output or background task can be re-directed to present their results at a subsequent time.”), 44:13-16 (“Another feature provided by the CVM 401 is conversational re-directions. As it is easy to re-

direct input and output of Unix processes, for example, conversational re-direction performs the same functions. However, the re-direction can be more sophisticated as it can involve partial mixed initiative notification while re-directing the streams.”.) Coffman describes examples of how the user can redirect system input/output, such as redirecting application output to use as input:

For example, with one conversational request (or by pressing a button on a keyboard or clicking a mouse or providing a key), the user can listen to the output of a task or re-claim the input (e.g. microphone) for a task down the context stack and direct a wave file, or an ASCII file, to append to the input stream. Similarly, the user can re-direct the printer where a file is sent, by giving a single redirecting request.

(Coffman, 35:2-6.)

134. The CVM kernel also mediates interaction between conversationally-aware applications and CVM subsystems, such as a speech recognition engine. (Coffman, e.g., 10:28-30 (“[T]he CVM provides conversational APIs and protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.”); 66:8-9 (“[T]he conversational kernel controls and accesses the conversational engines through a conversational engine API.”).) By default, the subsystem is “hidden to the application through the conversational application APIs.” (Coffman, 5:1-6.) But

Coffman explains that this default functionality can be overridden, such that an application can directly interface with the subsystem:

The conversational system 400 further comprises a plurality of conversational resource subsystems (engines) 407....[E]ach of these conversational subsystems 407 may be accessed through API calls to the CVM 401. The CVM 401 will locate the requested conversational subsystem 407 (via the conversational subsystem services manager 406), drive its execution and return appropriately the results. It is to be appreciated that these conversational subsystem 407 can be local or distributed over a network and that all conversational subsystem calls are hidden to the application (although the engine APIs are always available to the application if the developer wants a [sic] to implement a specific behavior of the engines 407).

(Coffman, 29:1-13.)

135. Turning last to the “**agent library**,” this is disclosed by Coffman’s conversational subsystems. This is because the conversational subsystems, consistent with 1[b][iii], “**include[] one or more utilities that can be used by the system agent and the plurality of domain agents**,” including speech recognition and text-to-speech utilities. (Coffman, 5:1-4 (“[T]he CVM utilizes conversational subsystems (which may be local or distributed) including speech recognition, speaker recognition, text-to-speech, natural language understanding and natural dialog generation engines to understand and generate dialog between and user and

machine.”.) With respect to the CVM software (“**system agent**”), Coffman discloses that it uses the utilities, explaining for example, that “[t]he CVM 401 will locate the requested conversational subsystem 407 (via the conversational subsystem services manager 406), drive its execution and return appropriately the results.” (Coffman, 29:7-9.) And conversationally-aware applications (“**plurality of domain agents**”), as explained, use the utilities either via CVM-mediated APIs or directly. (Coffman, 29:9-13.)

136. Coffman further teaches, and a person of ordinary skill in the art would have found it obvious, that the selected conversationally-aware application (“**the selected domain agent**”) uses the services of these components of the CVM and conversational subsystems (“**the communicatively coupled services**”) “**to create a response to the formulated question or command and format the response for presentation to the user.**”

137. The selected conversationally-aware application (“**the selected domain agent**”) uses the services of these components of the CVM and conversational subsystems (“**the communicatively coupled services**”) “**to create a response to the formulated question or command**” in a few different ways. First, these services enable the conversationally-aware application to register with the system in the first place, so that the application can be available to be selected to

respond to the user. As discussed, Coffman explains that a CVM includes a “conversational application & behavior/service manager layer 215” that “comprises functions for managing the conventional and conversationally aware applications 200 and 201,” including “keeping track of which applications are registered” and “the state of each application” (“**agent manager**”). (Coffman, 19:26-20:2.)

Second, as discussed, Coffman explains that the system’s conversational subsystems (“**agent library**”) include text-to-speech services to generate a spoken response to the user’s input based on a text response from the application. (Coffman, 10:28-30 (“In general, the CVM provides conversational APIs and protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.”), 29:1-4 (“The conversational system 400 further comprises a plurality of conversational resource subsystems (engines) 407 including, for example,...a text-to-speech synthesis (TTS) system 411”), 13:8-13 & 58:9-11 (CVM converts application “output into speech...provided to the user”), 44:21-22 (“[W]ith one conversational request...the user can listen to the output of a task”), 40:9-19, 13:8-13, 58:9-11, 39:3-5, Fig. 4 (showing TTS 411), Fig. 6 (showing speech synthesis 607 and “speech out” at telephone 600).) Either of these responses—the application’s text-based response or the generated spoken response based on the text response—would qualify as the

claimed “**response.**” The application’s text-based response would qualify because, as explained, the services of these components of the CVM and conversational subsystems enable the conversationally-aware application to register with the system in the first place. The spoken response would qualify because the application would use text-to-speech services in order to generate the response.

138. The selected conversationally-aware application (“**the selected domain agent**”) also uses the services of these components of the CVM and conversational subsystems (“**the communicatively coupled services**”) to “**format the response for presentation to the user.**” Coffman explains that the CVM kernel (“**system agent**”) generally provides “output formatting and presentation.” (Coffman, 20:8-14 (listing examples of “conversational services and behavior provided by the CVM kernel” including “output formatting and presentation”); *see also id.*, 19:26-20:27.) The CVM kernel is also “responsible for executing all the different functions needed to adapt the UI to the capabilities and constraints of the device, application and/or user preferences.” (Coffman, 20:15-18.) Coffman explains that one way in which the UI can be adapted is by using “voice fonts,” which “allows the user to select the type of voice providing the output.” (Coffman, 8:1-3; *see also id.*, 47:8-11 (“Another feature that is provided by the CVM is ‘conversational customization,’ whereby access to each task or resource can be

individually customized to preferences of the user requester. For instance, the personality/behavior of the CVM (e.g. synthesized voice - Voice Fonts) can be automatically customized to an identified user's preferences."), 39:30-32 ("It is to be appreciated that all the output, in particular voice output, can be customized and programmed by the user. Selection of the voice speaking the output can be made like fonts can be selected for text display. In such case, we speak of Voice fonts.").) Coffman accordingly make clear that one way in which the CVM kernel provides "output formatting and presentation" to "**format the response for presentation to the user**" is by using "voice fonts."

139. In addition, where the claimed "response" corresponds to the application's text-based response, text-to-speech provides an additional way in which the services are used to "**format the response for presentation to the user.**"

2. Claim 2: "The mobile system according to claim 1, wherein the natural language speech processing system further includes an event manager that:" (2[pre])

140. The written description of the '720 generally describes an "**event manager**" as a software module for managing events. ('720, 3:20-28 ("One aspect of the invention includes software that is installed onto the computer, where the software includes one or more of the following modules: ... an event manager for managing events and other modules.").) The '720 patent states, for example, that

“[a]gents **106** may receive and return events to the event manager **100.**” (’720, 21:61-62.) The ’720 patent’s use of the term “**event**” is consistent with its ordinary understanding in the art as broadly referring to an action or occurrence to which a program might respond. (*See, e.g.*, EX1009, p.198 (defining “event” as “[a]n action or occurrence, often generated by the user, to which a program might respond—for example, key presses, button clicks, or mouse movements”).)

141. A person of ordinary skill in the art would have understood and found it obvious that Coffman’s PVA-based conversational system (“**the mobile system**”) would include the claimed “**event manager.**” As explained below, the “**event manager**” in Coffman corresponds to software functionality that coordinates interaction among components of its “**natural language speech processing system**” comprising the conversational system’s CVM, conversational subsystems, and conversational APIs (application program interfaces) and protocols.

- (a) “**sends and receives events to a plurality of components of the natural language speech processing system to coordinate interaction among the plurality of components of the natural language speech processing system; and**” (2[a])

142. As discussed for 1[b], in Coffman, the claimed “**natural language speech processing system**” comprises the conversational system’s CVM, conversational subsystems, and conversational APIs (application program

interfaces) and protocols that enable the CVM, subsystems and conversationally-aware applications to communicate with each other. The “**event manager**” in Coffman corresponds to software functionality that coordinates interaction among these components, and includes functionality provided by “task dispatcher/controller 402” and “conversational subsystem services 406,” depicted in Figure 4 of Coffman:

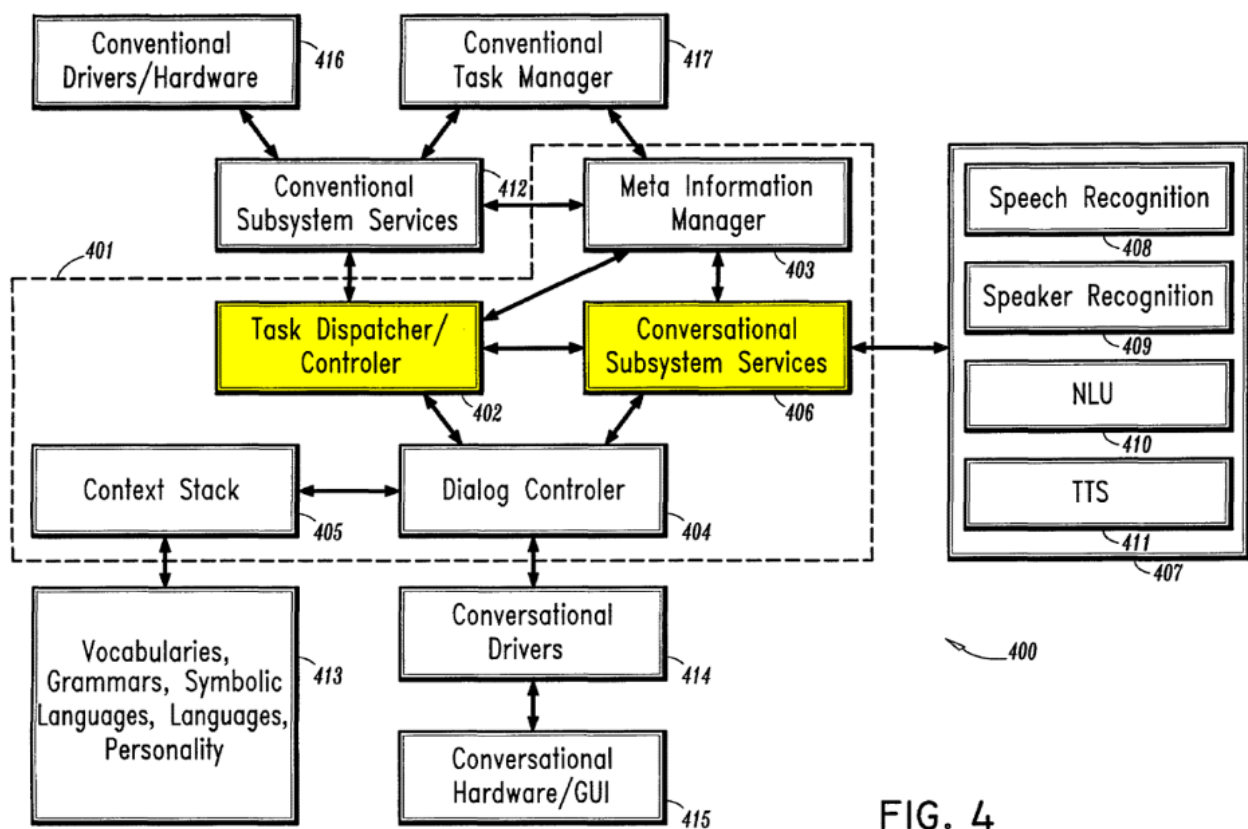


FIG. 4

(Coffman, Fig. 4 (highlighting added).)

143. Coffman makes clear that task dispatcher/controller 402 and conversational subsystem services 406 (“**event manager**”) “**coordinate interaction among**” the conversational system’s CVM, conversational subsystems, and

conversational APIs (application program interfaces) and protocols (“**a plurality of components of the natural language speech processing system**”). With respect to task dispatcher/controller 402, Coffman explains, for example, that “[b]ased on the modality (pointer, keyboard, file, speech), the task dispatcher 402 redirects the stream to the appropriate conversational subsystems or conventional subsystem with speech inputs being transcribed/understood” and that it also “feeds input from the conventional and conversational subsystems services 412, 406 to the context stack 405 (via the dialog controller 404 which selects the active context) and feeds the output of the different tasks to the different subsystems and prioritizes them.” (Coffman, 41:2-4, 31:15-18; *see also id.*, 30:26-28 (“The task dispatcher/controller 402 dispatches and coordinates different tasks and processes that are spawned (by the user and machine) on local and networked conventional and conversational resources.”).) And with respect to conversational subsystem services 406, Coffman explains that “[t]he CVM 401 will locate the requested conversational subsystem 407 (via the conversational subsystem services manager 406), drive its execution and return appropriately the results.” (Coffman, 29:7-9; *see also id.*, 29:14-15 (“The conversational subsystem services manager 406 manages all the services, UI and behavior (as described herein) that are offered by the CVM 401.”).)

144. A person of ordinary skill in the art would have also understood and

found it obvious that task dispatcher/controller 402 and conversational subsystem services 406 (“**event manager**”), in order to coordinate interaction, would “**send[] and receive[] events to**” the conversational system’s CVM, conversational subsystems, and conversational APIs (application program interfaces) and protocols (“**a plurality of components of the natural language speech processing system**”).

As noted, “**events**” broadly refer to actions or occurrences to which a program might respond. (*See, e.g.*, EX1009, p.198 (defining “event” as “[a]n action or occurrence, often generated by the user, to which a program might respond—for example, key presses, button clicks, or mouse movements”).) As discussed above, Coffman describes how task dispatcher/controller 402 and conversational subsystem services 406 receive and direct input and output (“**events**”) among the various CVM system components. (Coffman, 29:7-9, 29:14-15, 31:15-18, 30:26-28, 41:2-4.)

(b) “**provides a multi-threaded environment that enables the natural language speech processing system to provide a plurality of real-time responses to a plurality of questions or commands in a plurality of different contexts across a plurality of overlapping or interleaved sessions with the user.**” (2[b])

145. Task dispatcher/controller 402 and conversational subsystem services 406 (“**event manager**”) in Coffman’s conversational system “**provide a multi-threaded environment.**” As discussed above, Coffman teaches that task dispatcher/controller 402, among other things, manages threads/processes for the

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

system, explaining for example that it “divides each command/process into multiple actions, starts the associated threads/processes with the appropriate priority and relates/inserts them within the context stack 405.” (Coffman, 40:30-32; *see also id.*, 21:1-3 (“The task dispatcher/manager 221 dispatches and coordinates different tasks and processes that are spawned (by the user and machine) on local and networked conventional and conversational resources (explained in further detail below).”), 30:26-32 (“The task dispatcher/controller 402 dispatches and coordinates different tasks and processes that are spawned....”).)

146. The ability of task dispatcher/controller 402 provide a “**multi-threaded environment**” by managing threads/processes enables a user of Coffman’s conversational system to simultaneously interact with multiple applications. This simultaneous interaction entails Coffman’s “**natural language speech processing system**” “**provid[ing] a plurality of real-time responses to a plurality of questions or commands.**” Coffman describes, for example, that tasks can complete simultaneously and that collisions on the output stream resulting from simultaneously processing multiple user questions or commands may occur, thus making clear that the multi-threaded environment enables the system to provide a plurality of responses to a plurality of questions or commands. (Coffman, 42:12-13 (“Simultaneously completed tasks result in contexts that are arranged under the

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

active context according to CVM priorities (e.g. FIFO or FILO).”), 39:5-8 (“However, according to the context stack 405 and task dispatcher/controller 402, the output of multiple processes can simultaneously collide on the same output stream (e.g. a same display in text mode or the speech synthesizer).”).) Coffman also makes clear that responses are “**real-time**.” In the context of the ’720 patent, “real-time” relates to a time frame allowing for natural interaction with a user, and does not require, for example, an immediate or instantaneous response. The ’720 patent describes, for instance, providing “real-time response performance required for a natural interaction with the user.” (’720, 30:47-50.) This is also consistent with the general understanding of “real-time” in computer systems as describing computing activities that “match the human perception of time” or that are in response to “situations as they occur.” (EX1009, p.441 (defining “real-time” as “Of, or relating to, a time frame imposed by external constraints. Real-time operations are those in which the machines activities match the human perception of time or those in which the computer operations proceed at the same rate as a physical or external process. Real-time operations are characteristic of aircraft guidance systems, transaction-processing systems, scientific applications, and other areas in which a computer must respond to situations as they occur (for example, animating a graphic in a flight simulator or making corrections based on measurements).”).)

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

Coffman's descriptions mirror the '720 patent and this general understanding, describing "[t]he task dispatcher handles the issue of maintaining the dialog flow and, therefore, minimizes any delay die [sic; due] to the network and CPU load. It will prioritize the CPU cycles and available network route and resources to guarantee that delays on the dialog are minimized to acceptable levels." (Coffman, 45:29-32.) Moreover, it would have been obvious to a person of ordinary skill in the art for Coffman's conversational system to provide responses in "**real-time**," since the point of the system is to enable a user to carry on an interactive voice-based dialog with a computing system. (Coffman, 11:10-12 ("A CVM (or operating system) based on the conversational computing paradigm described herein according to the present invention allows a computer or any other interactive device to converse with a user.").) Providing only responses after an unacceptable delay would frustrate the conversational objectives of the system. For instance, Coffman provides an example command from a user to open a user interface window. (Coffman, 6:32-7:3 ("For example, a mouse click or pointer/stylus pointing action followed by the command 'I would like to open this' is disambiguated into a set of attribute value pairs: Command: open, Object: Windows or task selected by the last mouse click.").) Coffman provides additional examples of a user interacting with the system to make a phone call, obtain a grocery list and get driving directions. (Coffman, 56:26-57:9.)

If the system responded to the user input after only an unreasonable delay, the usefulness of Coffman's system would be significantly impaired.

147. Coffman also teaches that the plurality of questions or commands, and their responses, would be **“in a plurality of different contexts across a plurality of overlapping or interleaved sessions with the user.”** As discussed for 1[b][ii][1], Coffman explains that the conversational system includes a “context stack” where the state and history of each active process, such as an application process, is tracked in its own “context.” (Coffman, 29:18-30:3, 30:4-10, 5:15-20, 30:13-20.) A context stack thus comprises **“a plurality of different contexts”** that correspond to various applications and related input (questions or commands) and responses. (Coffman, e.g., 30:6-8 (“Each context associated with a given discourse comprises the latest requests made to the corresponding process/task/thread as well as the latest output.”), 30:13-25.) And the context stack enables the user to carry on **“overlapping or interleaved”** interactions with those various applications, where the user's ongoing interactions with each application is a different **“session”** (**“across a plurality of overlapping or interleaved sessions with the user”**).

3. **Claim 3: “The mobile system according to claim 2, wherein the multi-threaded environment of the event manager further enables the natural language speech processing system-to use a plurality of the domain agents to provide the plurality of real-time responses to the plurality of questions or commands in the plurality of different contexts across the plurality of overlapping or interleaved sessions with the user.”**

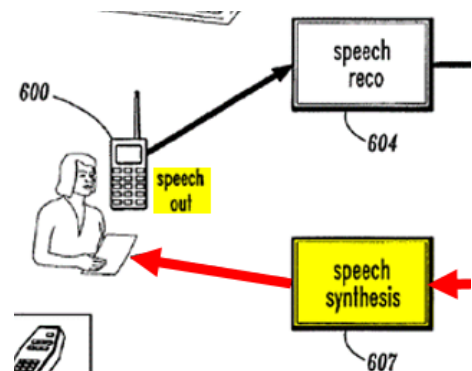
148. I addressed the language of claim 3 in my analysis of claim 2[b] above. As explained, Coffman’s task dispatcher/controller 402 (“**event manager**”) provides a “**multi-threaded environment**” by managing threads/processes that enables a user of Coffman’s conversational system to simultaneously interact with multiple applications. (Coffman, 21:1-3, 30:26-32, 40:30-32.) The applications provide “**real-time responses**” to the user’s questions or commands. (*E.g.*, Coffman, 45:29-32, 42:12-13, 39:5-8.) And with a context stack and its context entries, the plurality of questions or commands (and their responses) would be “**in a plurality of different contexts** across a plurality of overlapping or interleaved sessions with the user.” (Coffman, 5:15-20, 29:18-30:3, 30:4-10, 30:13-25.)

149. Claim 3 further adds that the claimed multi-threaded environment enables the natural language speech processing system “**to use a plurality of the domain agents**” to provide the plurality of real-time responses. The use of a plurality of domain agents in this manner was already addressed in my analysis of 2[b]. As explained, task dispatcher/controller 402 enables a user of Coffman’s

conversational system to simultaneously interact with multiple applications—i.e., **“to use a plurality of the domain agents.”** As noted, Coffman describes, for example, that tasks can complete simultaneously and that collisions on the output stream resulting from simultaneously processing multiple user questions or commands by multiple applications. (Coffman, 42:12-13 (“Simultaneously completed tasks result in contexts that are arranged under the active context according to CVM priorities (e.g. FIFO or FILO).”), 39:5-8 (“However, according to the context stack 405 and task dispatcher/controller 402, the output of multiple processes can simultaneously collide on the same output stream (e.g. a same display in text mode or the speech synthesizer).”).)

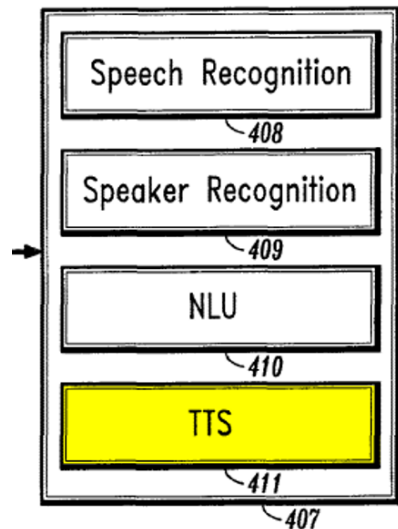
4. Claim 4: “The mobile system according to claim 1, wherein the response includes a text string and the natural language speech processing system further includes a text to speech engine that creates an encoded speech message to be annunciated to the user.”

150. I addressed the language of claim 4 in my analysis of 1[b][iii] above. As explained, Coffman discloses that a conversationally-aware application may provide a text-based response (**“the response include[] a text string”**) that is used to generate a spoken response to annunciate to the user created using a text-to-speech engine (**“a text to speech**



engine that creates an encoded speech message to be annunciated to the user”).

(Coffman, 10:28-30 (“In general, the CVM provides conversational APIs and



protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.”), 29:1-4 (“The conversational system 400 further comprises a plurality of conversational resource subsystems (engines) 407 including, for example,...a text-to-speech synthesis

(TTS) system 411”), Fig. 4 (excerpted at left, showing TTS 411), Fig. 6 (excerpt at right, showing speech synthesis 607 and “speech out” at telephone 600), 13:8-13 & 58:9-11 (CVM converts application “output into speech...provided to the user”), 44:21-22 (“[W]ith one conversational request...the user can listen to the output of a task”), 40:9-19, 13:8-13, 58:9-11, 39:3-5.) A person of ordinary skill in the art would have understood that the output from the text-to-speech engine would be an “**encoded**” speech message. (Coffman, 45:21-22 (describing “audio RecoVC (recognition compatible VoCoder) providing a coding protocol with pitch that allows reconstruction for play back etc.), 26:9-27 (describing speech coders).) This is because in a speech-processing system using conventional digital computing components, such as Coffman’s conversational system, a person of ordinary skill in

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

the art would have understood that, in order to process and output audio, the audio would be encoded in digital form. (Coffman, 13:1-3 (“The CVM system may be built on top of a conventional OS and APIs 15 and conventional device hardware 16 and located on a server or any client device (PC, PDA, Pvc).”); EX1009, pp.27 (“defining analog-to-digital converter” as “[a] device that converts a continuously varying (analog) signal, such as sound or voltage, from a monitoring instrument into binary code for use by a computer.”), 159 (defining “digital signal” as “[a] signal, such as one transmitted within or between computers, in which information is represented by discrete states—for example, high and low voltages—rather than by fluctuating levels in a continuous stream, as in an analog signal.”), 160 (defining “digital-to-analog converter” as “[a] device that converts digital data to an analog signal”), 338 (definition of “microphone” describing analog electrical signal provided by microphone being “convert[ed]...into digital data that a computer can process.”), 106 (defining “codec” as “1. Short for coder/decoder. Hardware that can convert audio or video signals between analog and digital forms.”).)

5. **Claim 5: “The mobile system according to claim 1, wherein the selected domain agent includes data associated with at least one of driving directions, travel information, restaurant information, systems information for the vehicle, safety information, or entertainment information.”**

151. A person of ordinary skill in the art would have understood and found

it obvious that a conversational application (“**the selected domain agent**”) in Coffman would include data associated with at least “**driving directions**” or “**travel information**.” Coffman provides an example where a conversational car PC (PVA) (“**mobile system**”) provides driving directions “tak[ing] into account the latest traffic information”:

You now instruct the PVA to guide you turn by turn to the store. Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information. At any time, you can request navigation information about where you are, what to do next, how far to go or even request a different itinerary.

(Coffman, 57:5-9.)

6. Claim 6: “The mobile system according to claim 1, wherein the selected domain agent includes data for communicating with one or more devices.”

152. A person of ordinary skill in the art would have understood and found it obvious that a conversationally-aware application (“the selected domain agent”) would “**include[] data for communicating with one or more devices.**”

153. For example, Coffman describes a situation in which a conversational car PC (PVA) provides driving directions to a store by fetching “the location of the store, locally or from a server.” (Coffman, 57:5-7 (“Your position is computed, the

location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information.”).) A person of ordinary skill in the art would have found it obvious that, in order for a navigation application to retrieve the location of the store from a server, the application would include data on how to communicate with the server, such as the server’s network address (e.g., an IP address), communication protocol (e.g., Hypertext transfer protocol (HTTP)) as well as required data and input format.

154. As another example, Coffman describes “a name dialer application [that] can be added to a conversational smart phone.” (Coffman, 45:2-3.) A name dialer application would include “**data for communicating with one or more devices**” in a variety of ways. A person of ordinary skill in the art would have understood that, in order for a name dialer application to instruct a phone to initiate a communication connection, the application would include data for how to communicate with the phone to carry out that instruction. Moreover, a person of ordinary skill in the art would have understood that the name dialer application would include phone numbers to dial. A phone number is “**data for communicating with one or more devices**” because it enables a communication connection to be established between devices. In addition, Coffman makes clear that a command to call someone using the application can involve the application directing a command

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

to another device, explaining that “unknown names or names that were never used before are sent to a more powerful networked machine for recognition and then download the updated information (phone number to dial etc.).” (Coffman, 45:2-6.) Like the navigation application retrieving location information from a server, a person of ordinary skill in the art would have understood that in order for the name dialer application to contact “a more powerful networked machine for recognition and then download the updated information (phone number to dial etc.),” the application would “**include data for communicating with**” that machine.

155. I further note that, with respect to the name dialer application, Coffman’s teachings, and the understanding of a person of ordinary skill in the art, are confirmed by Maes [EX1007], which Coffman incorporates by reference and which claims priority to the same United States provisional patent applications as Coffman, serial nos. 60/102,957 and 60/117,595. (Coffman, 22:25-29 (“patent application IBM Docket No. YO999- 113P, filed concurrently herewith, entitled ‘System and Method For Providing Network Coordinated Conversational Services,’ which is commonly assigned and incorporated herein by reference (wherein the conversational protocols are utilized in a system that does not utilize a CVM system).”), 1:2-3; *see also* Maes 7:12-16 (“the patent application IBM Docket No. YO999-111P, filed concurrently herewith, entitled ‘Conversational Computing Via

Conversational Virtual Machine,' which is commonly assigned and incorporated herein by reference.”), 1:4-5.) Maes further details regarding the operation of a name dialer application:

For example, assume the client device 100 is a smartphone with a name dialer local application. The user will store locally a list of desired names and addresses in the electronic address book of the smartphone. The user can then utter a command such as “dial *first name last name* at ... possible qualifier (home, office, cell phone), and upon recognition/understanding of the command (via the local conversational engines 102), the smartphone will automatically dial the phone number associated with the person in the address book (via the local applications 104). On the other hand, when a name is uttered that is not within the address book (and therefore not recognized/understood), but which is in a larger corporate (or public) directory (as contained in remote server 106), the request can be saved (in features or in waveform) and transmitted to a remote server 106 for recognition.

(Maes, 21:6-15.)

7. **Claim 7: “The mobile system according to claim 6, wherein the data for communicating with the one or more devices includes data for controlling the one or more devices.”**

156. I addressed the limitations of claim 7 in my analysis of claim 6. This is because, as explained, Coffman teaches that a conversationally-aware application,

such as a navigation application or name dialer application would include data to control another device so that it would operate in a particular way. For the navigation application, as discussed, it would control a remote server by instructing the server to retrieve and provide store location information. Similarly, for the name dialer application, it would control another machine by instructing that machine to provide contact information. The name dialer application would also control another device, such as a smart phone, by causing that device to initiate a connection request or causing another device to ring or otherwise respond to a connection request.

8. Claim 8: “The mobile system according to claim 6, wherein at least one of the one or more devices is connected to the vehicle.”

157. Claim 8 simply requires that one of the devices “**is connected**” to the vehicle. It does not require, for example, that the device be physically coupled to the vehicle. For example, the ’720 patent repeatedly describes wireless “connections.” (’720, 3:3-5 (describing “wireless connection” to vehicle device), 3:47-49, 5:38-42.) In a situation in which Coffman’s conversational system takes the form of a conversational car PC (PVA), it would have been obvious for one of the devices to be “connected” to the vehicle. For example, with respect to a navigation application running on the PVA in the vehicle, by contacting a remote server for information, the server would be “**connected to the vehicle.**” Similarly,

a name dialer application causing a connection to a remote machine or another phone device would result in that machine or other device being “**connected to the vehicle.**”

158. Moreover, even if claim 8 required that a device have some sort of physical connection to the vehicle, this also would have been obvious over Coffman. For example, as discussed for claim 6, a person of ordinary skill in the art would have understood that, in order for a name dialer application to instruct a phone to initiate a communication connection, the application would include data for how to communicate with the phone to carry out that instruction. Coffman provides an example in which both a conversational car PC (PVA) and a smart phone are in a user’s vehicle, and PVA uses the smart phone to dial a user’s desktop PC at home:

It does not matter if your desktop PC runs a CVM since you have, in your car, a conversational PDA, a conversational car PC (PVA, Personal Vehicle Assistant) and a conversational smart phone. Further assume that the PVA runs an embedded CVM while the two other applications are conversationally aware, i.e., you can control them through the CVM running on the PVA. You can instruct the PVA to dial in your PC using the phone. Once the connection is established, you are authenticated by voice and you find by voice the memo by simply requesting the “grocery list” that you had previously created, without having to remember the file name or the directory or browse through your directory to eventually select the appropriate file.

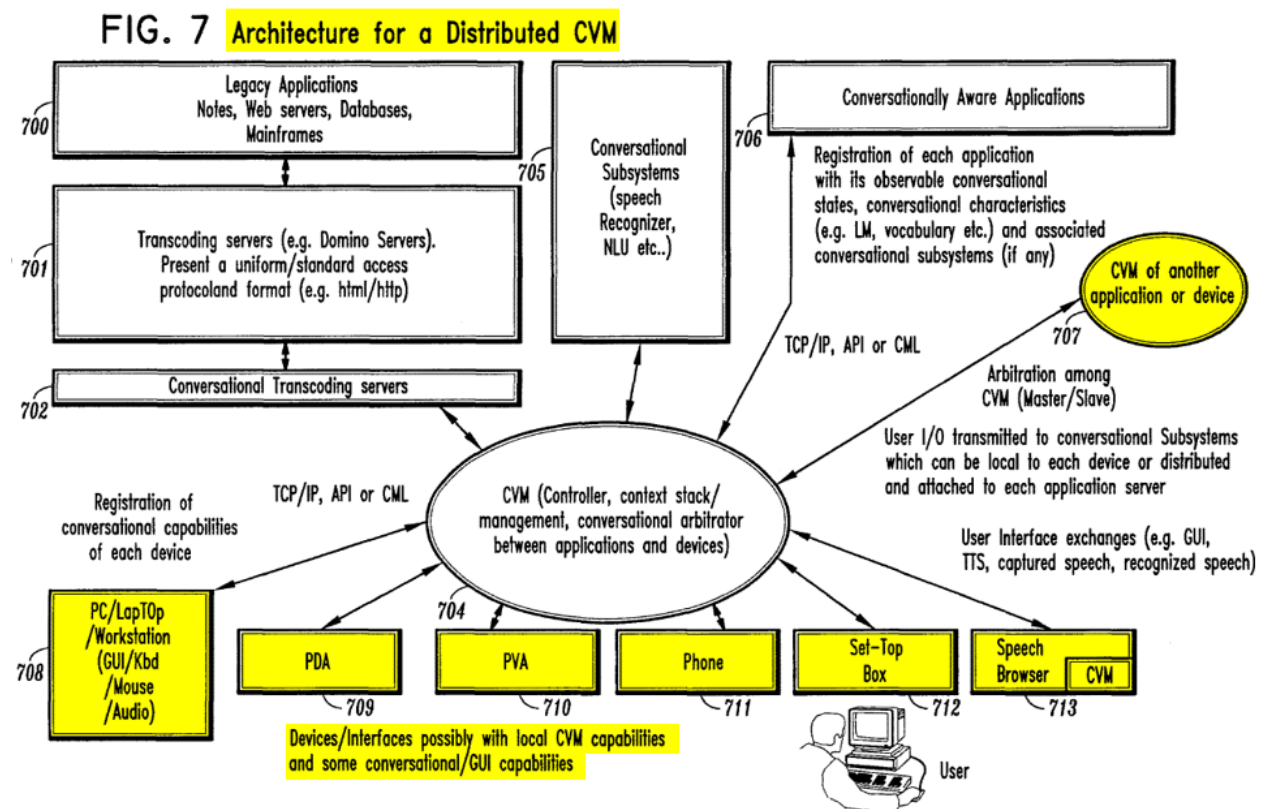
(Coffman, 56:22-30.) In the scenario outlined above, a person of ordinary skill in the art would have understood and found it obvious for the smart phone to be the claimed **“at least one of the one or more devices” “connected to the vehicle.”** It was routine and conventional to physically connect a phone to a vehicle, such as with an electrical connection or cradle/dock. (EX1009, p.245 (entry for “hands-free kit” describing “[w]ireless phone accessory” with “[m]ore elaborate sets for use in automobiles that may include a power amplifier, dashboard microphone, phone cradle, and speakers”); EX1017, pp.123 (defining “car phone” as “[t]he type of cellular phone that’s installed in a vehicle” and stating that “[a] car phone (also called a mobile unit) is attached to the vehicle, its power comes from the vehicle’s alternator (or battery if the car is not running)...”), 711 (definition of “transmobile” describing portable cellular phone that “draws its power from the vehicle’s battery via a cigarette lighter plug”), 136 (definition of “cellular radio” describing “dialing with the handset in the cradle”).)

9. Claim 9: “The mobile system according to claim 8, wherein at least one of the speech unit or the natural language speech processing system is located remotely from the vehicle and connected to the computer device on the vehicle through a communication link.”

159. It would have been obvious for both Coffman’s speech unit as well as natural language speech processing system to be **“located remotely from the**

vehicle and connected to the computer device on the vehicle through a communications link.”

160. As noted for 1 [b], Coffman discloses that the conversational system can be a distributed system involving multiple devices connected via communication links, as illustrated in Figure 7:



(Coffman, Fig. 7 (highlighting added).) As shown in the figure above, the distributed architecture can involve multiple devices running CVMs. Coffman provides an example architecture where both a conversational car PC and a user’s desktop PC are running CVMs. (Coffman, 57:22-58:6 (“With CVM running on the

desktop and on the PVA...”).) In such an architecture, the CVM at the user’s desktop PC can be designated as the master CVM. (Coffman, 58:4-6 (“When the PVA conversationally connects to the PC, it will be up to the application developer of the PVA, to decide if the PVA acts as master, slave or partner.”).)

161. Therefore, when a desktop PC is acting as a master CVM for example, Coffman makes clear that the PC could include a “**speech unit**” as described for 1[a] enabling a user to use their voice to control a conversational car PC, explaining:

In addition, the distributed conversational protocols 300 also allow the applications and devices to exchange other information such as applets, ActiveX components, and other executable code that allows the devices or associated applications to coordinate a conversation between such devices in, e.g., a master/slave or peer-to-peer conversational network configuration and networks comprising silent partners. In other words, when multiple CVM or conversationally aware multiple devices are conversationally connected and coordinated, it becomes possible to simultaneously control them through one single interface (i.e. through a single microphone). For example, voice can automatically control through a unique coordinated conversational interface a smart phone, a pager, a PDA, networked computers, a IVR and a car embedded computer.

(Coffman, 23:16-25.) In this arrangement, the “**speech unit**” would be “**located remotely from the vehicle and connected to the computer device on the vehicle**

through a communication link.”

162. Similarly, Coffman contemplates that, with a desktop PC acting as the master CVM for example, voice input at the conversational car PC could be transmitted to the desktop PC for processing by the desktop PC’s CVM and associated conversational subsystems (“**natural language speech processing system**”). (Coffman, 25:27-26:1 (“More specifically, multiple CVM devices when registering will add to the conversational registration capability, information pertaining to, e.g., their controlling capability, the conversational engines that they have access to, and applications and devices that have registered with them and that they control. Based on their UI, I/O capabilities and active I/O, one CVM controller becomes the master and the other CVM controllers act as slaves, which is equivalent relatively to the master as being registered applications until a new negotiation occurs.”).) For example, Coffman describes “speech transmission protocols 302 (or conversational coding protocols)” that are “used by speech transmission clients 38, 38a to transmit/received compressed speech to/from other networked devices, systems or applications for processing.” (Coffman, 26:4-6; *see also id.*, 26:6-25 (describing speech transmission protocols).) Therefore, using these protocols, voice input to a speech unit at a conversational car PC could be transmitted to a desktop PC conversational system for processing, and the results returned to the

conversational car PC. In such an arrangement, the **“natural language speech processing system”** (at the desktop PC) would be **“located remotely from the vehicle and connected to the computer device on the vehicle through a communication link.”**

- 10. Claim 10: “The mobile system according to claim 8, wherein the at least one device connected to the vehicle includes at least one of a navigation system, a monitoring system, a security system, a control system, or a media system connected to the vehicle.”**

163. As discussed for claims 6 and 8, Coffman describes connecting to a device providing a navigation service to the vehicle (**“the at least one device connected to the vehicle”**). (Coffman, 57:5-7 (“Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information.”).) Therefore, Coffman discloses **“wherein the at least one device connected to the vehicle includes...a navigation system.”** Therefore, claim 10 is obvious.

- 11. Claim 11: “The mobile system according to claim 8, further comprising a telematics control unit that interfaces with the at least one device connected to the vehicle.”**

164. Consistent with the ordinary understanding of **“telematics,”** the '720 patent acknowledges that “telematics systems” broadly encompass “systems that bring human-computer interfaces to vehicular environments.” ('720, 1:25-26;

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

EX1009, p.514 (defining “telematics” as “[i]n communications technology, the linking of computers and telecommunications. Telematics technology is becoming standard in the automotive industry, with dashboard navigation systems, roadside assistance, entertainment, Internet, and cellular services available in vehicles.”).) A “**telematics control unit**” simply refers to a device that controls telematics functionality. (E.g., EX1009, p.128 (defining “control unit” as “[a] device or circuit that performs an arbitrating or regulating function. For example, a memory controller chip controls access to a computer’s memory and is the control unit for that memory.”).) By 2001-2002, “**telematics control unit[s]**” were well-known and in commercial use, and thus neither inventive nor non-obvious. (EX1009, p.514 (definition of “telematics” stating that “[t]elematics technology is becoming standard in the automotive industry”); EX1017, pp.685-686 (definition of “telematics” stating that “[p]robably the best known example is GM’s OnStar system, which automatically calls for assistance if the vehicle is in an accident.”).)

165. In Coffman, the claimed “**telematics control unit**” corresponds to the vehicle’s conversational car PC (PVA). As discussed, Coffman explains how the PVA serves as a central unit through which a user interfaces with the CVM system to access functionality provided by the system’s applications. (E.g., Coffman, 55:27-28 (“When a user interacts with the collection of devices, the interaction may

always be via a central unit such as a PVA (personal vehicle assistant) 710 in a car,
or a speech browser 713.”), 56:24-25 (“Further assume that the PVA runs an
embedded CVM while the two other applications are conversationally aware, i.e.,
you can control them through the CVM running on the PVA.”).) Coffman discloses
that, through the PVA, the user is able to access various telematics functions, such
as telephony and navigation. (Coffman, 56:24-57:16 (“Further assume that the PVA
runs an embedded CVM while the two other applications are conversationally aware,
i.e., you can control them through the CVM running on the PVA. You can instruct
the PVA to dial in your PC using the phone....You now instruct the PVA to guide
you turn by turn to the store....Pressed by time, you instruct the PVA to dial the store
drive-through server.”).)

166. As explained for claim 8, Coffman discloses that the vehicle’s
conversational car PC (PVA) (“**telematics control unit**”) is also connected to the
user’s smart phone in the vehicle as well as servers providing functionality for
navigation and name dialing applications. As explained, each of these devices
discloses the claimed “**at least one device connected to the vehicle.**” Because the
PVA communicatively connects to each of these devices, Coffman discloses that the
PVA (“**telematics control unit**”) “**interfaces with the at least one device
connected to the vehicle.**” (*E.g.*, Coffman, 55:27-28, 56:24-25.)

12. Claim 12: “The mobile system according to claim 11, wherein at least one of the speech unit or the natural language speech processing system is embedded in the telematics control unit.”

167. As explained for claim 1, a person of ordinary skill in the art would have understood and found it obvious that Coffman’s conversational car PC (PVA) (“**telematics control unit**”) would include a “**speech unit**” and/or a “**natural language speech processing system**.” (E.g., Coffman, 56:21-25 (“[Y]ou have, in your car, a conversational PDA, a conversational car PC (PVA, Personal Vehicle Assistant) and a conversational smart phone. Further assume that the PVA runs an embedded CVM while the two other applications are conversationally aware, i.e., you can control them through the CVM running on the PVA.”), 55:27-28 (“When a user interacts with the collection of devices, the interaction may always be via a central unit such as a PVA (personal vehicle assistant) 710 in a car, or a speech browser 713.”).) Because the PVA serves as the central unit for a user to interact with the CVM system and can comprise conventional computer software and hardware, it would have been obvious for the “**speech unit**” (e.g., microphone and speakers) as well as the “**natural language processing system**” to be “**embedded**” in the PVA (“**telematics control unit**”). (Coffman, 13:1-3 (“The CVM system may be built on top of a conventional OS and APIs 15 and conventional device hardware 16 and located on a server or any client device (PC, PDA, Pvc).”), 29:1-13

(conversational subsystem can be local), 55:27-28, 56:21-25.)

13. Claim 13: “The mobile system according to claim 1, wherein the communicatively coupled services include at least one remotely located service and the selected domain agent includes data for controlling or communicating with the remotely located service.”

168. As discussed for claim 6, Coffman provides examples of conversationally-aware applications that communicate with a remotely located service, and it would have been obvious for the applications to include data for controlling or communicating with that service. For example, Coffman teaches a navigation application that communicates with a remote server that provides a location/navigation service. (Coffman, 57:5-7 (“Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information.”).) As another example, Coffman also teaches a name dialer application that communicates with remote machine providing a contact information service. (Coffman, 42:2-6 (“[U]nknown names or names that were never used before are sent to a more powerful networked machine for recognition and then download the updated information (phone number to dial etc.).”.) More generally, Coffman describes how a CVM device “can access desired information from a service network provider network 1001 by connecting via a PSTN 1002 and internet/intranet 1003 networks through router 1004.” (Coffman,

63:3-6 (“In the system depicted in Fig. 12, client devices 1000 (equipped with CVM system or dialog manager capabilities) according to the present invention can access desired information from a service network provider network 1001 by connecting via a PSTN 1002 and internet/intranet 1003 networks through router 1004.”).)

169. Therefore, it would have been obvious for “**the communicatively coupled services to include at least one remotely located service**” and for a conversationally-aware application (“**the selected domain agent**”) to “**include[] data for controlling or communicating with the remotely located service.**”

14. Claim 14: “The mobile system according to claim 13, wherein the remotely located service includes at least one of a payment service provider, a customer relationship management system, a specialized service, a location service, or an emergency service.”

170. As discussed for claim 6, Coffman teaches a remotely located service that includes “**a location service.**” (Coffman, 57:5-7 (“Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information.”).)

15. Claim 15: “The mobile system according to claim 13, wherein the speech unit and the natural language speech processing system each include respective transceivers that communicate via a communication network.”

171. As discussed for claim 9, Coffman contemplates that, with a desktop PC acting as the master CVM, voice input (via a “**speech unit**”) at the conversational

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

car PC could be transmitted to the desktop PC for processing by the desktop PC's CVM and associated conversational subsystems ("**natural language speech processing system**"). (Coffman, 25:27-26:1.) For example, Coffman describes "speech transmission protocols 302 (or conversational coding protocols)" that are "used by speech transmission clients 38, 38a to transmit/received compressed speech to/from other networked devices, systems or applications for processing." (Coffman, 26:4-6; *see also id.*, 26:6-25 (describing speech transmission protocols).) A person of ordinary skill in the art would have understood that this arrangement would have predictably resulted in the "**speech unit**" and the "**natural language speech processing system**" of Coffman's conversational system "**communicat[ing] via a communication network**," such as a public switched telephone network (PSTN) and internet. Coffman describes, for example, how "client devices 1000 (equipped with CVM system or dialog manager capabilities) according to the present invention can access desired information from a service network provider network 1001 by connecting via a PSTN 1002 and internet/intranet 1003 networks through router 1004." (Coffman, 63:3-6.) It would have been obvious, therefore, that a conversational car PC (PVA), acting as a client of a master CVM at the user's desktop PC, would communicate with it via a communication network.

172. A person of ordinary skill in the art would have also understood and

found it obvious that the “**speech unit**” of the conversational car PC and the “**natural language processing system**” of the desktop PC would “**each include respective transceivers**” for communicating with each other. A “**transceiver**” is a conventional and generic device for transmitting and receiving data. (EX1009, p.527 (defining “transceiver” as “[s]hort for **transmitter/receiver**. A device that can both transmit and receive signals. On LANs (local area networks), a transceiver is the device that connects a computer to the network and that converts signals to and from parallel and serial form.”); *see also* EX1017, p.709 (defining “transceiver” as “[a]ny device that transmits and receives”).) It therefore would have been obvious that, in order for the “**speech unit**” and “**natural language processing system**” to communicate, each would include a “**transceiver.**”

16. Claim 16: “The mobile system according to claim 15, wherein the communication network is a wide area wireless network.”

173. It would have been obvious that the communication network between the speech unit and the natural language processing system in Coffman’s conversational computing system could be a “**wide area wireless network.**” As discussed for claim 15, Coffman describes communicating via PSTN and internet, which is an example of a “**wide area network.**” (Coffman, 63:3-6; EX1009, p.284 (definition of “internetwork” describing communication over “the Internet or

another WAN (wide-area network)").) And because the communication involves a device in a mobile vehicle, a person of ordinary skill in the art would have found it obvious that the communication network would be a "**wide area wireless network**," such as connecting to the internet via a cellular modem. (EX1017, p.135 (defining "cellular modem" as "[a] device the combines a data modem and cellular telephone transceiver technologies in a single unit").) Coffman's mere passing mention of "wireless network" in a list of various types of networks confirms that communication via wireless networks was well-known and conventional by 2001-2002. (Coffman, 63:14 ("In any network (internet, bluetooth, wireless network etc..) such as shown in Fig. 12...").)

17. Claim 22: "The mobile system according to claim 1, wherein the speech unit is located remotely from the natural language speech processing system."

174. For purposes of my analysis, claim 22 is substantially the same as claim 9, which similarly recites "[t]he mobile system according to claim 8, wherein at least one of the speech unit or the natural language speech processing system is located remotely from the vehicle and connected to the computer device on the vehicle through a communication link." As discussed for claim 9, it would have been obvious for the "**speech unit**" and "**natural language processing system**" to be located remotely from each other, such as a speech unit in a user's vehicle and a

natural language processing system at the user's home desktop PC.

18. Claim 23: “The mobile system according to claim 1, further comprising at least one of a display or a keypad.”

175. It would have been obvious for Coffman's conversational car PC (PVA)-based conversational system to comprise “**at least one of a display or a keypad.**” Coffman describes a “multi-modal” system and makes clear that the system is not limited to devices lacking a display or keypad. In fact, Coffman explains that “[i]ndeed, a computer with classical keyboard inputs and pointing devices coupled with traditional monitor display can profit significantly by utilizing the CVM according to the present invention.” (Coffman, 11:29-30.) Coffman explains that “[t]he ‘multi-modality’ aspect of the CUI [conversational user interface] implies that various I/O resources such as voice, keyboard, pen, and pointing device (mouse), keypads, touch screens, etc can be used in conjunction with the CVM platform.” (Coffman, 12:8-10.) Including a display and/or keypad with a PVA would beneficially provide the user with additional ways of inputting and receiving information. For example, as discussed, Coffman describes a scenario involving a user receiving turn by turn guidance for driving to a store. (Coffman, 57:5-9.) Consistent with conventional in-vehicle navigation systems at the time, a person of ordinary skill in the art would have readily appreciated that a PVA with a display to show a map to the user would likely improve the user experience by

providing further information about the user's circumstances. (*E.g.*, EX1009, p.514 (definition of "telematics" describing "dashboard navigation systems").)

19. **Claim 24: "The mobile system according to claim 1, wherein at least one of the speech unit or the natural language speech processing system is embedded in at least one of the vehicle, a handheld device connected to the vehicle, a fixed computer device connected to the vehicle, or a mobile computer device connected to the vehicle."**

176. For purposes of my analysis, claim 24 is substantially the same as claim 12, which similarly recites "[t]he mobile system according to claim 11, wherein at least one of the speech unit or the natural language speech processing system is embedded in the telematics control unit." As discussed for claim 12, it would have been obvious for the "speech unit" and/or "natural language speech processing system" to be "embedded" in Coffman's conversational car PC (PVA), which Coffman teaches could in turn be "embedded in...the vehicle" or "a fixed device connected to the vehicle." (Coffman, 12:23-25 ("a car embedded computer"), 23:23-25 (same), 55:27-28 ("a central unit such as a PVA (personal vehicle assistant) 710 in a car").)

20. **Claim 25: "The mobile system according to claim 1, wherein the communicatively coupled services include at least one shared network resource."**

177. It would have been obvious for "the communicatively coupled services" to "include at least one shared network resource." As discussed for

claim 6, Coffman provides examples of conversationally-aware applications that communicate with network-based services such as a location service and a contact information service. (Coffman, 57:5-7, 42:2-6.) It would have been obvious for the resources provided by these services (such as databases storing location and contact information) to be shared among devices and/or users in order to provide more useful and widely-accessible services. This is confirmed by the Maes application incorporated by reference into Coffman [EX1007], which further describes a name dialer application, explaining that the contact information service can be provided by “a larger corporate (or public) directory.” (Maes, 21:12-15.)

21. Claim 27: “The mobile system according to claim 1, wherein the one or more utilities included in the agent library include at least one of a text and string handling, function, a network communications function, a database lookup and management function, a fuzzy and probabilistic evaluation function, or a text to speech formatting function.”

178. The utilities in Coffman’s conversational subsystems (“**agent library**”) include at least “**text and string handling**,” “**fuzzy and probabilistic evaluation**” and “**text to speech formats**.”

179. With respect to “**text and string handling**,” a person of ordinary skill in the art would have understood that Coffman’s conversational subsystems perform text and string handling in numerous ways. For example, Coffman explains that “using the conversational engines 208 and conversational data files 209 (accessed

by CVM 202 via conversation engine APIs 207), each input is converted into a binary or ASCII input, which can be directly processed by the programming language as built-in objects.”⁵ (Coffman, 18:27-30; *see also id.*, 6:26-28, 52:24-27.) Coffman further explains that “[e]ach input stream is converted into its ASCII transcription and aligned with input time-marks via speech recognition processing 604.” (Coffman, 38:9-11.) Coffman also explains that “Queries can be decoded, parsed and then translated into a logic combination of queries (symbolic query) using NLU technology.” (Coffman, 51:12-14.) A person of ordinary skill in the art would have understood that these transcription, parsing and translating operations would involve “text and string handling” functions.

180. With respect to “**fuzzy and probabilistic evaluation**,” a person of ordinary skill in the art would have understood that Coffman’s speech recognition process (performed by the speech recognition engine) would involve a fuzzy and probabilistic evaluation of the spoken input using a language model to identify the most likely words comprising the input. As explained for claim 1, under the proposed combination with Kanevsky, Coffman’s speech recognition would have involved using trigram language model probabilities to determine the most likely

⁵ “ASCII input” refers to text strings formatted using the American Standard Code for Information Interchange (“ASCII”). (EX1009, p.36 (definition of “ASCII”).)

sequence of words in a user's utterance. The fuzzy/probabilistic evaluation is reflected in the fact that the determination does not reflect certainty but rather a probability of being correct.

181. With respect to “**text to speech formats**,” Coffman, as noted, identifies text-to-speech synthesis as one of its conversational subsystems. (Coffman, 29:1-4.) Coffman makes clear that text-to-speech synthesis involves using “voice fonts” that allow the speech to be formatted in a particular type of voice. (Coffman, 47:8-13 (“For instance, the personality/behavior of the CVM (e.g. synthesized voice - Voice Fonts) can be automatically customized to an identified user's preferences.”), 8:1-3 (“advanced conversational interfaces...with multiple personalities, such as Voice fonts, which allows the user to select the type of voice providing the output.”), 39:30-32 (“[A]ll the output, in particular voice output, can be customized and programmed by the user. Selection of the voice speaking the output can be made like fonts can be selected for text display. In such case, we speak of Voice fonts.”).)

22. Claim 29: “The mobile system according to claim 1, wherein the speech recognition engine:” (29[pre])

182. As discussed, Coffman discloses a conversational computing system involving a PVA (“**mobile system**”). The system includes speech processing functionality, including speech and speaker recognition, that collectively corresponds to the “**speech recognition engine.**” (Coffman, Fig. 4 (“Speech

Recognition 408” & “Speaker Recognition 409”), 29:1-4.) As explained below, it would have been obvious for Coffman’s “**speech recognition engine**” to perform the following actions.

(a) “determines an identity of the user based on unique voice characteristics of the user; and” (29[a])

183. Coffman describes using “text-independent speaker identification” to identify a user. (Coffman, 47:25-27 (“In a speech-based system, each command can be used to perform text-independent speaker identification.”), 8:16-18 (“A user is automatically identified whenever a query to a resource is made.”), 46:9-12 (“[W]ith speech-based conversational systems, since each command conveys not only the formulation of the query but also enough information for authentication of the user using, [sic] text-independent speaker verification can be used to identify and verify a user.”).) “Text-independent speaker identification,” as its name suggests, is a form of speaker identification using some features of a speech signal to determine the identity of the speaker of a given utterance that does not require the utterance to comprise specific linguistic content. (See Gibbon [EX1008], pp.474 (defining “speaker recognition” as “[a]ny decision-making process that uses some features of the speech signal to determine some information on the identity of the speaker of a given utterance”), 486 (defining “text-independent speaker recognition system” as “[a] speaker recognition system for which the linguistic content of test speech

utterance varies across trials”).) The very fact that text-independent speaker identification is *text-independent* implies that the identification does not rely on specific words but instead uses the user’s unique voice characteristics. Coffman therefore discloses and renders obvious “**determin[ing] an identity of the user based on unique voice characteristics of the user.**” Kanevsky, which as noted, Coffman incorporates by reference (Coffman, 37:5-13), provides further confirmation, explaining that text independent speaker recognition that involves comparing “**voice characteristics**” of a sample of the user’s utterance with a model to verify the user’s identity. (Kanevsky, e.g., 6:66-7:8, 6:12-15 (“The identity claim may be made from other information provided by the speaker or voice characteristics...”), 2:6-8 (describing prior art systems that “employ voice-based verification” using “acoustic characteristics of the utterances spoken by the user”).)

(b) “**tags the recognized words or phrases with the identity of the user to associate the utterance with the user and the current dialog in the natural language speech processing system.**” (29[b])

184. “Tagging” is generally understood in the art as creating an association between a piece of information and another piece of information, such as a characteristic or property. The written description of the ’720 patent does not discuss “tagging” in detail, but consistent with the ordinary understanding, states that “the tags...may allow other components of the system to tie [an] utterance to the correct

user and dialog.” (’720, 27:5-12.)

185. A person of ordinary skill in the art would have understood and found it obvious that Coffman’s “**speech recognition engine**,” which both recognizes words and phrases in a user’s utterance and determines the user’s identity, would “**tag**” the recognized words and phrases with the user identity. Coffman makes clear that an association is created between the recognized words and phrases and the user identity, explaining that “user identity can be associated in the context of each discourse,” so that, for example, “command ambiguity can be immediately and transparently resolved (e-mail from my mother is correctly understood, independently of the user).” (Coffman, 48:1-3; *see also id.*, 46:26-31 (“[T]he stack of contexts should contain the identity of the user as the most recently authenticated identity”), 48:11-14 (“[W]ith respect to conversational multi-users and conversational security, it is to be appreciated that dialogs, categories, meta-information, and access to resources can be a function of the identity of the user and its associated meta-information history. And conversely, the conversational information collected on a query can be used to recognize the user.”).) A person of ordinary skill in the art would have therefore understood and found it obvious that speaker identity associated with the recognized words and phrases as well as with context/discourse would be used to determine a context and discourse for a current

dialog that best match the user's recognized words and phrases. A person of ordinary skill in the art would have accordingly understood and found it obvious that such a matching process would involve creating an association between ("**tag[ging]**") "**the recognized words or phrases with the identity of the user**" "**to associate the utterance with the user and the current dialog in the natural language speech processing system.**"

23. **Claim 30: "The mobile, system according to claim 1, wherein the plurality of dictionary and phrase entries used by the speech recognition engine are further dynamically updated based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the one or more prior dialogs."**

186. My analysis for 1[b][i] above fully addressed claim 30. As explained, "**the plurality of dictionary and phrase entries**" under the proposed combination is disclosed by a trigram language model provided by a conversationally-aware application to a conversational system and used for speech recognition. The step of "**dynamically updat[ing]**" the entries "**based on one or more dynamic fuzzy set possibilities or prior probabilities derived from the history of the current dialog and the one or more prior dialogs**" simply corresponds to dynamically updating the trigram language model and its prior probabilities in accordance with model adaptation as taught by Kneser as the user interacts (via dialog) with the conversational system.

24. Independent Claim 31: “A method responsive to a user generated natural language speech utterance, comprising:” (31[pre])

187. For purposes of my analysis, the preamble of claim 31 is substantially the same as the preamble of claim 1. The preamble of claim 31 recites “**a method,**” instead of claim 1’s “mobile system,” “**responsive to a user generated natural language speech utterance.**” As explained below, the claimed method is disclosed and rendered obvious by the prior art.

(a) “receiving, at a speech unit connected to a computer device on a vehicle, a natural language speech utterance from a user, wherein the speech unit converts the received natural language speech utterance into an electronic signal;” (31[a])

188. 31[a] simply recites a method counterpart to 1[a], which similarly recites “a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user and converts the received natural language speech utterance into an electronic signal.” Therefore, for the reasons discussed for 1[a], 31[a] is disclosed and obvious.

- (b) **“recognizing, at a speech recognition engine connected to the computer device on the vehicle, at least one of words or phrases from the electronic signal, wherein the speech recognition engine uses at least data received from a plurality of domain agents to recognize the words or phrases, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;” (31[b])**

189. 31[b] is largely a method counterpart 1[b][i], which similarly recites “a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user.”

190. 31[b] further recites that the speech recognition engine is **“connected to the computer device on the vehicle.”** This was addressed in my analysis of claim 1[b], which explained that the claimed “natural language speech processing system” (of which the speech recognition engine is a part) would be connected to the conversational car PC (PVA) (**“connected to the computer device on the vehicle”**).

191. Therefore, for the reasons discussed for 1[b] and 1[b][i], 31[b] is

disclosed and obvious.

(c) **“determining, at a parser connected to the computer device on the vehicle, a context for the natural language speech utterance;” (31[c])**

192. 31[c] is largely a method counterpart to 1[b][ii] and 1[b][ii][1], which similarly recite “a parser” (1[b][ii]) that “determin[es] a context for the natural language speech utterance” (1[b][ii][1]).

193. 31[c] further recites that the parser is **“connected to the computer device on the vehicle.”** This was addressed in my analysis of claim 1[b], which explained that the claimed “natural language speech processing system” (of which the parser is a part) would be connected to the conversational car PC (PVA) (**“connected to the computer device on the vehicle”**).

194. Therefore, for the reasons discussed for 1[b], 1[b][ii] and 1[b][ii][1], 31[c] is disclosed and obvious.

(d) **“selecting, at the parser connected to the computer device on the vehicle, at least one of the plurality of domain agents based on the determined context;” (31[d])**

195. 31[d] is largely a method counterpart to 1[b][ii] and 1[b][ii][2], which similarly recite “a parser” (1[b][ii]) that “select[s] at least one of the plurality of domain agents based on the determined context” (1[b][ii][2]).

196. Like 31[c], 31[d] also recites that the parser is **“connected to the**

computer device on the vehicle,” which as explained, was addressed in my analysis of claim 1[b].

197. Therefore, for the reasons discussed for 1[b], 1[b][ii] and 1[b][ii][2], 31[d] is disclosed and obvious.

- (e) **“transforming, at the parser connected to the computer device on the vehicle, the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command; and” (31[e])**

198. 31[e] is largely a method counterpart to 1[b][ii] and 1[b][ii][3], which similarly recite “a parser” (1[b][ii]) that “transform[s] the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command” (1[b][ii][3]).

199. Like 31[c], 31[e] also recites that the parser is **“connected to the computer device on the vehicle,”** which as explained, was addressed in my analysis of claim 1[b].

200. Therefore, for the reasons discussed for 1[b], 1[b][ii] and 1[b][ii][3], 31[e] is disclosed and obvious.

- (f) **“forwarding the formulated question or command to an agent architecture connected to the computer device on the vehicle, wherein the agent architecture communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.” (31[f])**

201. 31[f] is largely a method counterpart to 1[b][iii], which similarly recites “an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.”

202. 31[f] further recites that the agent architecture is **“connected to the computer device on the vehicle.”** This was addressed in my analysis of claim 1[b], which explained that the claimed “natural language speech processing system” (of which the agent architecture is a part) would be connected to the conversational car PC (PVA) (**“connected to the computer device on the vehicle”**).

203. Unlike 1[b][iii], 31[f] further recites **“forwarding the formulated**

question or command” to the agent architecture. But this also would have been obvious. As explained for 1 [b][iii], the “**agent architecture**” in Coffman comprises the conversational APIs and protocols that enable communication between the CVM, conversational subsystems, and conversationally-aware applications. The APIs and protocols enable the CVM and conversational subsystems to communicate with conversationally-aware applications. (Coffman, e.g., 10:28-30 (“In general, the CVM provides conversational APIs and protocols between conversational subsystems (e.g. speech recognition engine, text-to speech etc.) and conversational and/or conventional applications.”), 5:1-6, 12:28-30, 13:17-28, 13:29-32, 19:16-19, 22:10-23:3, 24:10-11, Fig. 2.) Therefore, it would have been obvious that an application call/query (“**the formulated question or command**”) would have been “**forwarded**” from the parser (where it was formulated) to the conversational APIs and protocols in order to transmit it to the selected conversationally-aware application for processing.

25. Claim 35: “The method according to claim 31, the wherein the selected domain, [sic] agent includes data for controlling or communicating with at least one of a navigation system, a monitoring system, a security system, a control system, or a media system connected to the vehicle.”

204. For purposes of my analysis, claim 35 is essentially an amalgam of claims 6, 8 and 10. As discussed for those claims, it would have been obvious for

Coffman's conversational car PC (PVA) to include a navigation application (“**selected domain agent**”) having data for controlling or communicating with a server-based navigation system communicatively connected to the vehicle.

26. Claim 36: “The method according to claim 31, wherein the selected domain agent includes data associated with at least one of driving directions, travel information, restaurant information, systems information for the vehicle, safety information, or entertainment information.”

205. For purposes of my analysis, claim 36 is not meaningfully different from claim 5, which similarly recites “[t]he mobile system according to claim 1, wherein the selected domain agent includes data associated with at least one of driving directions, travel information, restaurant information, systems information for the vehicle, safety information, or entertainment information.” As discussed for claim 5, a person of ordinary skill in the art would have understood and found it obvious that a conversational application (“**the selected domain agent**”) in Coffman would include data associated with at least “**driving directions**” or “**travel information.**”

27. Claim 37: “The method according to claim 31, wherein the communicatively coupled services include at least one remotely located service and the selected domain agent includes data for controlling or communicating with the remotely located service.”

206. For purposes of my analysis, claim 37 is not meaningfully different

from claim 13, which similarly recites “[t]he mobile system according to claim 1, wherein the communicatively coupled services include at least one remotely located service and the selected domain agent includes data for controlling or communicating with the remotely located service.” Claim 37 therefore would have been obvious for the same reasons discussed for claim 13.

28. Claim 38: “The method according to claim 37, wherein forwarding the formulated question or command to the agent architecture includes transmitting a request to the remotely located service.”

207. It would have been obvious to a person of ordinary skill in the art that **“forwarding the formulated question or command to the agent architecture”** would include **“transmitting a request to the remotely located service.”** As discussed for claim 13, Coffman provides examples of conversationally-aware applications that communicate with a remotely located service, such as a location/navigation service. (*E.g.*, Coffman, 57:5-7 (“Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information.”).) More generally, Coffman describes how a CVM device “can access desired information from a service network provider network 1001 by connecting via a PSTN 1002 and internet/intranet 1003 networks through router 1004.” (Coffman, 63:3-6 (“In the system depicted in Fig. 12, client devices 1000 (equipped with CVM system or

dialog manager capabilities) according to the present invention can access desired information from a service network provider network 1001 by connecting via a PSTN 1002 and internet/intranet 1003 networks through router 1004.”.) Therefore, it would have been obvious that forwarding, for example, an application call/query for turn-by-turn directions to a store (“**the formulated question or command**”) to Coffman’s agent architecture for processing by a conversationally aware application would “**include transmitting a request**” as claimed because it would entail sending a request to the remotely located location/navigation service to “fetch” the store’s location.

29. Claim 39: “The method according to claim 38, wherein the remotely located service is associated with a remotely located device.”

208. Claim 39 was largely addressed in my analysis of claim 38. A person of ordinary skill in the art would have understood that a remotely located service, such as a location/navigation service described by Coffman, would be “**associated with a remotely located device**” such as the server providing that service. (Coffman, 57:5-7, 63:3-6.) The location/navigation service would be “**associated**” with that server because, simply put, the server provides the service and moreover, a person of ordinary skill in the art would have understood that information associating the service with the server (such as an IP address or URL) would be used

in order for a CVM device to be able to communicate with the remotely located service provided by the server. Therefore, claim 39 would have been obvious.

30. Claim 40: “The method according to claim 38, wherein the request is transmitted to the remotely located service via a communication network.”

209. As discussed for claim 38, Coffman describes how a CVM device can communicate with a remotely located service via a communication network, such as “a PSTN 1002 and internet/intranet 1003 networks.” (Coffman, 63:3-6.) It therefore would have been obvious that a PVA’s communication with, for example, the remotely located location/navigation service described in Coffman would take place “**via a communication network.**”

31. Claim 41: “The method according to claim 37, wherein the remotely located service includes at least one of a payment service provider, a customer relationship management system, a specialized service, a location service, or an emergency service.”

210. For purposes of my analysis, claim 41 is not meaningfully different from claim 14, which similarly recites “[t]he mobile system according to claim 13, wherein the remotely located service includes at least one of a payment service provider, a customer relationship management system, a specialized service, a location service, or an emergency service.” As discussed for claim 14, Coffman

teaches a remotely located service that includes “**a location service.**” (Coffman, 57:5-7 (“Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information.”).)

32. Claim 42: “The method according to 38, wherein the request is transmitted to the remotely located service via a wide-area RF transceiver.”

211. As discussed for claim 38, Coffman describes transmitting a request from a vehicle to a remotely located service to retrieve information about a store’s location in order to provide turn by turn directions to the store (“**the request [] transmitted to the remotely located service**”). (Coffman, 57:5-7.) It would have been obvious that such a request would have been transmitted “**via a wide-area RF transceiver,**” such as a cellular transceiver. This is because it would have involved communicating with a remote service from a highly mobile environment, a situation routinely addressed by using cellular devices and networks. For example, Coffman describes using a smart phone, a type of cellular device, in a user’s car in order to communicate with the user’s remotely located PC. (Coffman, 56:26 (“You can instruct the PVA to dial in your PC using the phone”); *see also* EX1009, p.485 (explaining that a “smart phone” is a “hybrid between a wireless telephone and a personal digital assistant (PDA)”), EX1017, p.135 (definition of “cellular radio”

describing cellular transceivers that each provide radio communication coverage over a wide area, such as “several city blocks”).)

33. Claim 43: “The method according to claim 31, wherein the communicatively coupled services include one or more shared network resources.”

212. For purposes of my analysis, claim 43 is not meaningfully different from claim 25, which similarly recites “[t]he mobile system according to claim 1, wherein the communicatively coupled services include at least one shared network resource.” Claim 43 therefore would have been obvious for the same reasons discussed for claim 25.

34. Claim 52: “The method according to claim 31, further comprising providing interactive guidance to assist the user in performing one or more maneuvers to operate the vehicle.”

213. As discussed, Coffman describes the PVA providing “turn by turn” guidance to a store, using the vehicle’s position, the location of the store, and “tak[ing] into account the latest traffic information.” (Coffman, 57:5-9 (“You now instruct the PVA to guide you turn by turn to the store. Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information. At any time, you can request navigation information about where you are, what to do next, how far to go or even request a different itinerary.”).) The turn-by-turn guidance is “**interactive**”

because it provides guidance as the user operates their vehicle in accordance with the directions, it takes into account traffic information, and the user is able to interact with it by requesting navigation information. (Coffman, 57:5-9.) A person of ordinary skill in the art would have understood that the turn-by-turn guidance also assists the user in performing one or more “**maneuvers**”—such as turns and lane changes—to operate the vehicle. Coffman therefore discloses providing “**interactive guidance**” (turn-by-turn guidance) “**to assist the user in performing one or more maneuvers to operate the vehicle**” (e.g., turns and lane changes).

35. Claim 55: “The method according to claim 31, further comprising providing interactive guidance to relating to one or more tasks for the user based on a location of the vehicle.”

214. As discussed, Coffman describes the PVA providing “turn by turn” guidance to a store, using the vehicle’s position, the location of the store, and “tak[ing] into account the latest traffic information.” (Coffman, 57:5-9 (“You now instruct the PVA to guide you turn by turn to the store. Your position is computed, the location of the store is fetched, locally or from a server, and an itinerary is computed to take into account the latest traffic information. At any time, you can request navigation information about where you are, what to do next, how far to go or even request a different itinerary.”).) The turn-by-turn guidance is “**interactive**” because it provides guidance as the user operates their vehicle in accordance with

the directions, it takes into account traffic information, and the user is able to interact with it by requesting navigation information. (Coffman, 57:5-9.) Coffman therefore discloses providing turn-by-turn guidance (“**interactive guidance**”) for driving to a store (“**relating to one or more tasks for the user**”) using the vehicle’s location (“**based on a location of the vehicle**”).

C. Ground 2: Claims 18, 20-21 and 32-33 Are Obvious Over Ground 1 Prior Art in Further View of DeLine (EX1010).

- 1. Claim 18: “The mobile system according to claim 1, wherein the speech unit includes a speech coder that converts the natural language speech utterance into the electronic signal, an array microphone that receives the natural language speech utterance, and a filter that optimizes a signal to noise ratio of the electronic signal.”**

215. Turning first to the “**speech coder**,” a “**coder**” is simply functionality that can convert an analog audio or video signal to a digital form. (See EX1009, p.106 (defining “codec” as “**1. Short for coder/decoder. Hardware that can convert audio or video signals between analog and digital forms.**”).) A person of ordinary skill in the art would have therefore understood and found it obvious that a user device in Coffman’s system, such as a PVA, would include a “**speech unit**” with functionality to convert a user’s speech into a digital signal—i.e., “**a speech coder**”—as claimed.

216. In fact, Coffman expressly discloses that its devices include a “**speech**

coder.” Referring to Figure 3 (shown below), which depicts a distributed conversational computing system comprising a client device and other networked devices, Coffman describes using “speech coders” and “acoustic signal processing engines...and suitable audio hardware” to carry out distributed speech processing:

The speech coders 234, 234a provide perceptually acceptable or intelligible reconstruction of the compressed speech and optimized conversational performance (e.g., word error rate). The speech is captured (and transformed into features) on the respective networked devices using acoustic signal processing engines (audio subsystems) 232, 232a and suitable audio hardware 233, 233a. In addition, compressed speech file formats 303 can be transmitted and received between devices for processing speech. More specifically, the speech transmission protocols 303 allow the devices to transmit and receive compressed speech or local processing results to/from other devices and applications on the network.

(Coffman, 26:9-16; *see also id.*, 26:16-27.)

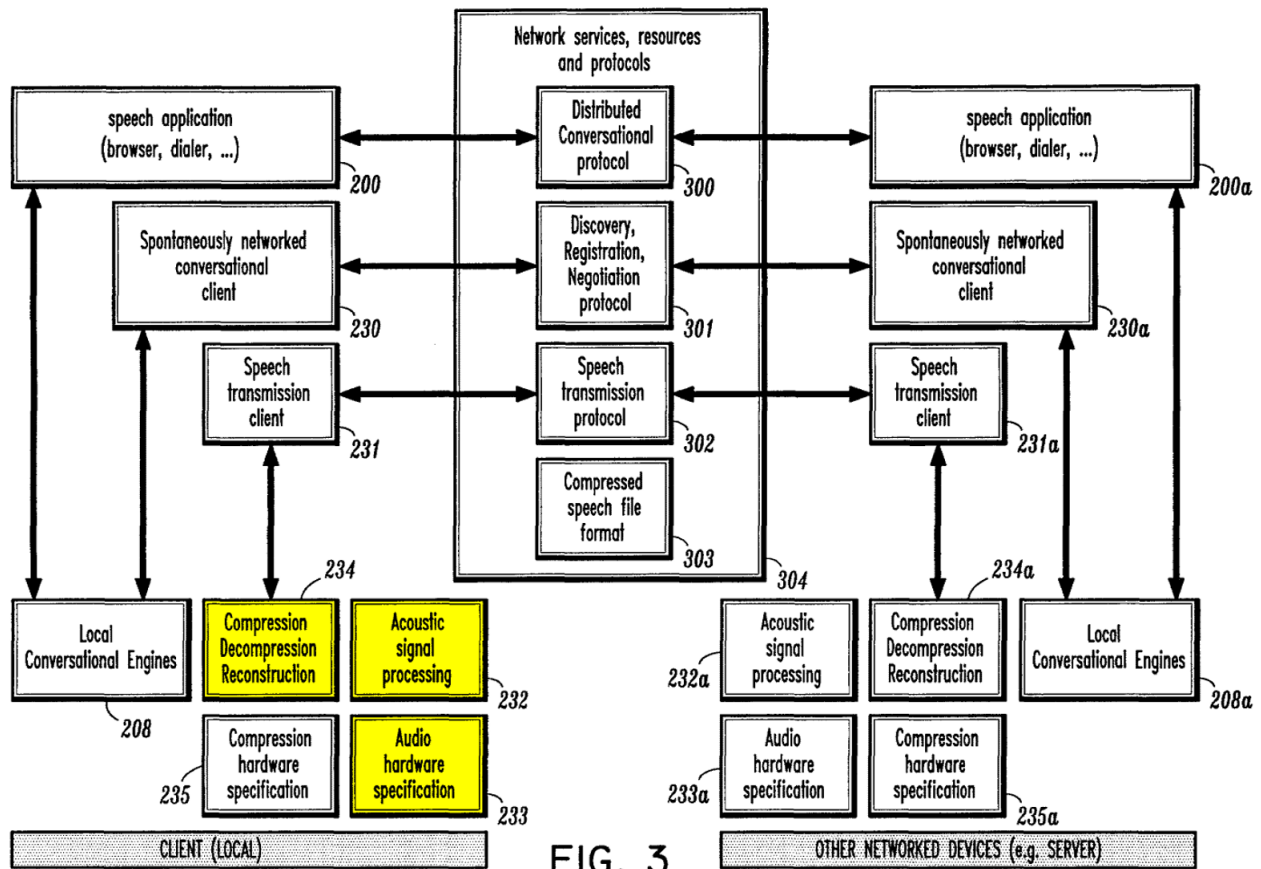


FIG. 3

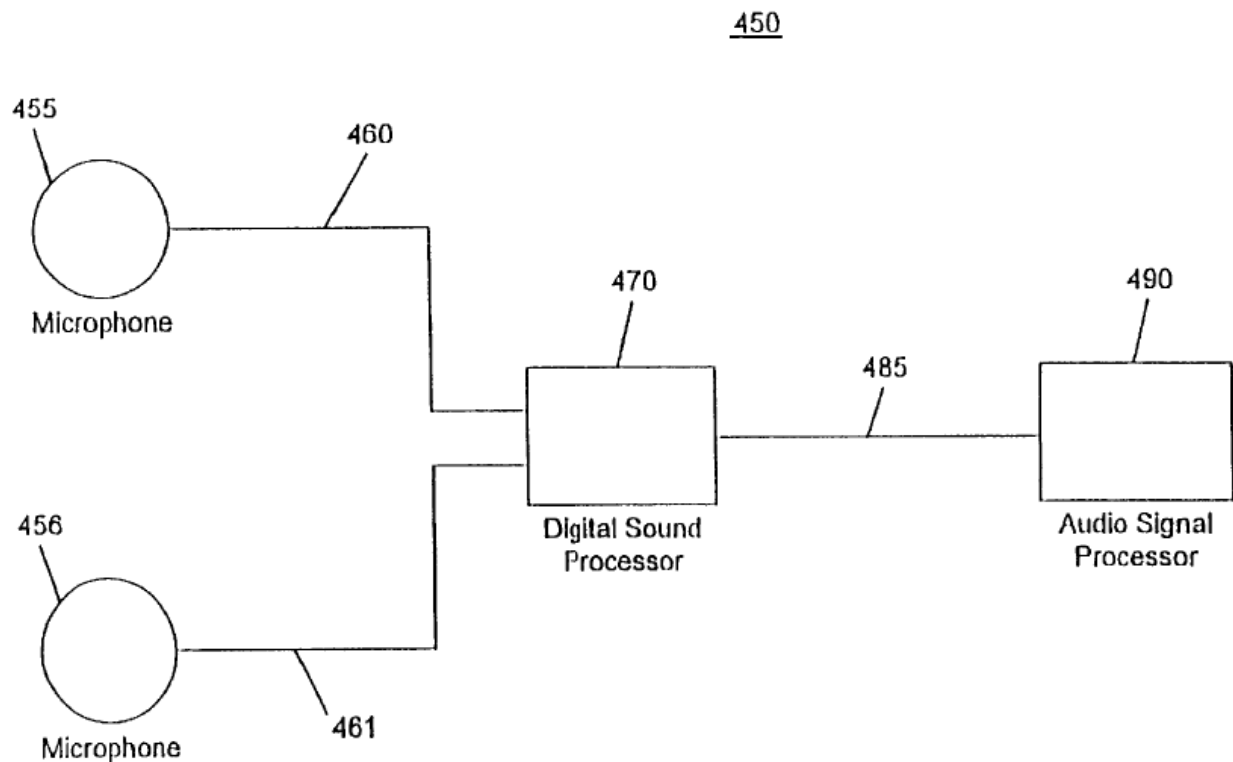
(Coffman, Fig. 3 (highlighting added), 23:3-5; *see also id.*, Fig. 7 (illustrating distributed architecture).) A person of ordinary skill in the art would have understood and found it obvious that a client device’s “speech coder” and/or “acoustic signal processing engines” (and “suitable audio hardware”) in Coffman’s conversational system would correspond to a **“speech coder that converts the natural language speech utterance into the electronic signal”** as claimed; and further, that as part of the functionality for processing incoming speech, it would comprise part of the device’s **“speech unit.”**

217. With respect to **“an array microphone that receives the natural language speech utterance,”** Coffman does not go into detail about microphone arrangements in its conversational system, instead briefly noting how “it becomes possible to simultaneously control [multiple devices] through one single interface (e.g., through a single microphone).” (Coffman, 7:15-18.) The same is true of Coffman for **“a filter that optimizes a signal to noise ratio of the electronic signal.”** As noted above, Coffman only notes “acoustic signal processing engines.” (Coffman, 26:11-13, Fig. 3.) But an **“array microphone”** and **“filter”** as claimed would have been obvious in view of U.S. Patent No. 6,420,975 B1 (DeLine [EX1010]).

218. DeLine, entitled “Interior Rearview Mirror Sound Processing System,” generally describes an in-vehicle microphone arrangement and sound processing system for enhancing vocal content and reducing ambient noise content detected by the microphone(s). (DeLine, 1:20-28.) DeLine explains that many vehicles use hands-free communication devices to avoid problems that may arise when a driver has to hold the device while driving. (DeLine, 1:29-32.) But with such devices, DeLine explains, the vocal content detected by the microphone may be degraded. (DeLine, 1:40-2:33.) For example, “ambient noise from the like of traffic noise, road noise, wind noise, HVAC noise and engine noise can make it difficult for the

remote party to voice recognize with clarity and accuracy the message being telecommunicated.” (DeLine, 2:29-33.) DeLine explains that “[t]his may be especially troublesome when the other party is an automated voice system which responds to a particular voice communication.” (DeLine, 2:18-20.) “When a remote receiving system is operating in voice recognition mode, it is particularly important that alphanumeric speech text is accurately received and interpreted by the remote party.” (DeLine, 2:20-23; *see generally id.*, 2:15-33.)

219. In one embodiment of its system, DeLine proposes a microphone array comprising multiple microphones, as illustrated in Figure 12 below:



(DeLine, Fig. 12; *see also id.*, 47:19-53, Fig. 14 (depicting two-microphone mirror-

mounted system comprising microphones **520** and **530**), 48:53-63 (describing “array of multiple mirror-mounted microphones” and commercially-available microphone array from Andrea Electronics Corporation).) “FIG. 12 describes a vehicular digital sound processing **450** utilizing two microphones **455**, **456** that provide signals **460**, **461**, respectively, as inputs to digital sound processor **470** which provides vocal signature signal **485** to audio vehicle signal processor **490**.” (DeLine, 47:19-23.) DeLine explains the benefits and advantages of using input from multiple microphones:

By receiving separate inputs (and with each individual input optionally filtered to enhance their principal signal content), digital sound processor **470** can mathematically/computationally analyse the respective inputs, and their combination, in order to extract, using noise cancellation techniques as known in the art, vocal signature signal **485** which is input to audio vehicle signal processor **490** (for example, an ONSTAR™ system) where signal **485** is clearly and accurately received and processed, including in voice recognition mode. Use of at least two microphones in vehicular digital sound processing is beneficial to facilitate generation of high quality vocal signals because destructive interference techniques and/or signal averaging techniques can be utilized to cancel out unwanted noise and to isolate the desired vocal signature signal that is indicative of the alphanumeric words spoken by the driver and/or other vehicle occupant.

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

(DeLine, 47:32-48.) As noted in the excerpt above, DeLine describes how the use of multiple microphones enables a digital sound processor to use “noise cancellation techniques as known in the art” to “facilitate generation of high quality vocal signals.” (*Id.*) “Such noise-cancelling techniques,” DeLine explains, “include techniques such as echo cancellation including echo cancellation algorithms; beam forming techniques including beamforming digital algorithms and adaptive beamforming; active noise cancellation; open and closed loop noise cancellation; adaptive signal processing techniques and adaptive noise cancellation techniques such that when, for example, speech by a driver is received at and detected by microphone 520, the output from the opposing microphone 530 is analyzed for ambient noise content (and vice versa should speech by a passenger be detected by microphone 530).” (DeLine, 48:26-45.)

220. Turning back to the claim language, an “**array microphone**” is a device comprising an arrangement of multiple microphones. DeLine’s use of two or more microphones in this manner discloses “**an array microphone that receives the natural language speech utterance.**” (*E.g.*, DeLine 48:53-63 (describing “array of multiple mirror-mounted microphones” and commercially-available microphone array from Andrea Electronics Corporation).) I note that, during prosecution of the ’720 patent, the Examiner cited DeLine for its disclosures related

to this limitation, which the applicant did not dispute. (EX1019, pp.00357-00359 (pending claims 18, 20 and 29 reciting “array microphone”), 00412-00413 (office action rejecting claims 18, 20 and 29 in view of DeLine), 00441-00442 (applicant response).)

221. With respect to the claimed “**filter that optimizes a signal to noise ratio of the electronic signal,**” this is disclosed by DeLine’s discussion of system components to perform various noise-cancelling techniques, as discussed above. (DeLine, 47:32-48, 48:26-45, 48:53-59 (describing “digitally filter[ing]” microphone output).) The noise-cancelling techniques performed by, e.g., digital sound processor **470**, serve as a “filter” to remove noise, such as ambient cabin noise, thereby optimizing the signal to noise ratio of the electronic signal from the microphones. (DeLine, e.g., 1:20-27 (system for “providing a vocal signal indicative of speech detected by the vehicular microphone with an enhanced vocal content and a reduced ambient noise content”).)

222. **Rationale and Motivation to Combine (Ground 1 Prior Art and DeLine)**. It would have been obvious to a person of ordinary skill in the art to combine DeLine with Coffman and its conversational computing system under the proposed combination, predictably resulting in the speech unit of a PVA in Coffman’s conversational system further including multiple microphones and

components for performing noise-cancelling operations on the electronic signals provided from the microphones. Combining DeLine and Coffman in this manner would have therefore resulted in a speech unit that included **“an array microphone that receives the natural language speech utterance”** as well as **“a filter that optimizes a signal to noise ratio of the electronic signal.”**

223. DeLine is an analogous reference in the same field of speech recognition and processing as the '720 patent. ('720, 1:12-21; DeLine, e.g., 1:20-34, 2:15-33.) DeLine would have also been reasonably pertinent to a number of problems facing the patent inventors, including how to improve the quality of speech signals received in a vehicle environment or other environment with ambient noise. (DeLine, 2:15-33.)

224. A person of ordinary skill in the art would have been motivated to combine Coffman and DeLine. As discussed, Coffman describes using its conversational computing system in a vehicle environment. DeLine teaches that, in such an environment, a user's speech input may be degraded, and that this “may be especially troublesome” when interacting with a voice-controlled system such as Coffman's. (DeLine, 2:15-33.) Because DeLine teaches that its techniques of using multiple microphones and various noise-cancelling filter techniques improve the voice signal and result in better speech processing and recognition, a person of

ordinary skill in the art would have been motivated to incorporate them into the similar environment contemplated by Coffman to obtain the same benefits and advantages (or at least obvious to try them in Coffman's system).

225. A person of ordinary skill in the art would have had a reasonable expectation of success. The incorporation into Coffman's system of multiple microphones and signal filtering as described by DeLine would have involved the application of conventional techniques and technologies. DeLine describes for example, how various noise-cancelling techniques are "known in the art," and in fact, identifies a commercially-available array microphone system that could be used. (DeLine, 47:32-41, 48:30-45, 48:53-63.) A person of ordinary skill in the art would have found the disclosures of Coffman and DeLine sufficiently detailed to implement the techniques on which I rely, without undue experimentation.

2. Claim 20: "The mobile system according to claim 18, wherein the array microphone is at least a one-dimensional array."

226. Claim 20 does not add anything that was not addressed in my analysis of claim 18 above. The '720 patent makes clear that a **"one-dimensional"** microphone array is simply a linear array. ('720, 25:48-53 ("Some embodiments of the invention may use one or more arrays of microphones **134** to provide better directional signal capture and noise elimination than can be achieved with a single

microphone. The microphone array can be one-dimensional (a linear array) or two-dimensional (a circle, square, triangle or other suitable shape).”).) An arrangement of two microphones (or more), as described in DeLine, would be “**at least a one-dimensional**” (or linear) array. As noted above, during prosecution of the ’720 patent, the applicant did not dispute that DeLine disclosed this limitation. (EX1019, pp.00357 (pending claim 20), 00412 (office action rejecting claim 20 in view of DeLine), 00441-00442 (applicant response).)

3. Claim 21: “The mobile system according to claim 18, wherein the speech coder uses an adaptive lossy audio compression to convert the natural language utterance into the electronic signal.”

227. As discussed above for claim 18, Coffman discloses a “**speech coder**” that uses “**audio compression**” for encoding a user’s utterance as an electronic signal. (Coffman, Fig. 3, 26:9-16 (“[T]he speech transmission protocols 303 allow the devices to transmit and receive compressed speech or local processing results to/from other devices and applications on the network”); *see also id.*, 26:6-9 (“The speech transmission clients 38, 38a operates in conjunction with compression, decompression and reconstruction engines 234, 234a using suitable compression hardware 235, 235a for processing the speech transmitted over the network.”), 26:13-14 (“In addition, compressed speech file formats 303 can be transmitted and received between devices for processing speech.”); *see generally id.*, 26:4-27.)

228. It would have been obvious for Coffman's speech coder to use **"adaptive lossy audio compression."** Adaptive lossy audio compression was a well known compression technique that a person of ordinary skill in the art would have recognized was particularly suited to processing and transmitting voice data in a mobile system. For example, a person of ordinary skill in the art would have understood that "lossy" audio compression beneficially reduced the amount of data needed to encode voice data and its use would thereby improve system processing and response times. (*E.g.*, EX1009, pp.319 (definition of "lossy compression" explaining that "[v]ideo and sound files often contain more information than is apparent to the viewer or listener; a lossy compression method, which does not preserve that excess information, can reduce data to as little as 5 percent of its original size."), 117 (definition of "compress" describing reducing size of a set of data "so that it can be stored in less space or transmitted with less bandwidth" and that "lossy compression can be used for video and sound files").) In fact, the obviousness of using **"adaptive lossy audio compression"** is confirmed by the Maes application that is incorporated by reference into Coffman [EX1007]. Coffman points to Maes for additional information about encoding speech. (Coffman, 26:23-24 ("The packet headers preferably specify the coding scheme and coding arguments (i.e. sampling frequency, feature characteristics, vector

dimensions, feature transformation/family etc. as discussed in the above incorporated patent application IBM Docket No. YO999-113P) using for encoding the speech (or results).”).) Maes makes clear that a person of ordinary skill in the art would have recognized that “any suitable compression scheme” for encoding speech could be used. (Maes, 16:5-6.) Maes then recognizes the importance of bit rate in optimizing compressed audio and specifically describes how, in order to adhere to a target bit rate, the compression rate “can be **adapted** between robustness (error rate - feature distortion) and perceptual quality.” (Maes, 16:9-15.) Maes therefore confirms that a person of ordinary skill in the art would have considered “**adaptive lossy audio compression**” to be a “suitable compression scheme” and obvious choice to use in Coffman’s system.

4. **Claim 32: “The method according to claim 31, wherein the speech unit includes an array microphone that receives the natural language speech utterance, a speech coder that converts the natural language speech utterance into the electronic signal, and a filter that optimizes a signal to noise ratio of the electronic signal.”**

229. For purposes of my analysis, claim 32 is substantially identical to claim 18, and therefore would have been obvious for the same reasons.

5. **Claim 33: “The method according to 32, wherein the filter removes background noise from the electronic signal to optimize the signal to noise ratio of the electronic signal.”**

230. Claim 33 does not add anything that was not addressed in my analysis

of claim 18 above. As discussed, DeLine discloses using various noise-cancelling techniques to optimize the signal to noise ratio of the electronic signals from the multiple microphones by removing ambient (background) noise. (DeLine, 47:32-41, 48:30-45, 1:20-27 (system for “providing a vocal signal indicative of speech detected by the vehicular microphone with an enhanced vocal content and a reduced ambient noise content”), 48:53-59 (describing “digitally filter[ing]” microphone output).) Therefore, for the reasons discussed for claim 18, it would have been obvious for the speech unit of a PVA in Coffman’s conversational system to include a filter that **“removes background noise from the electronic signal to optimize the signal to noise ratio of the electronic signal.”**

D. Ground 3: Claims 19 and 34 Are Obvious Over Ground 2 Prior Art in Further View of Gerson (EX1011).

6. Claim 19: “The mobile system according to claim 18, wherein the filter employs adaptive echo cancellation to optimize the signal to noise ratio of the electronic signal.”

231. As discussed for claim 18, it would have been obvious in view of DeLine for the speech unit of a PVA in Coffman’s conversational computing system to include a filter providing various noise-cancelling techniques. (DeLine, 48:30-45.) With respect to claim 19, DeLine identifies “echo cancellation including echo cancellation algorithms” as one of those techniques. (DeLine, 48:33-36.) DeLine does not explicitly state that echo cancellation techniques are “adaptive.” However,

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

DeLine also identifies various “adaptive signal processing techniques and adaptive noise cancellation techniques” (DeLine, 48:38-40), which suggests that the echo cancellation would include adaptive techniques. To the extent there is any question that DeLine discloses or suggests “adaptive echo cancellation,” it would have been obvious in further view of U.S. Patent No. 6,937,977 B2 (**Gerson** [EX1011]).

232. Gerson, entitled “Method and Apparatus for Processing an Input Speech Signal During Presentation of an Output Audio Signal,” “relates generally to communication system incorporating speech recognition and, in particular, to a method and apparatus for ‘barge-in’ processing of an input speech signal during presentation of an output audio signal.” (Gerson, 1:8-12.) Gerson explains that “[g]iven human nature and the design of some speech synthesis/recognition systems, the spoken responses provided by a caller will often occur during presentation of an output audio signal, for example, a synthesized speech prompt.” (Gerson, 1:35-39.) “The processing of such occurrences is often referred to as ‘barge-in’ processing.” (Gerson, 1:39-40.) In a “barge-in” situation, “the speech recognition system must account for residual artifacts from the prompt being present in any spoken response provided by the user (i.e., an input speech signal) in order to effectively perform speech recognition.” (Gerson, 1:44-50.)

233. Gerson’s system includes an “echo cancellation and environmental

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

processing (ECEP) block” that receives input from a microphone and “provides echo cancellation of the speaker audio **211** from the microphone audio **220**.” (Gerson, 9:5-11, 9:23-26.) Gerson explains that “[t]his form of echo cancellation is known as acoustic echo cancellation and is well known in the art” and cites U.S. Patent No. 5,561,668 as an example (which it incorporates by reference). (Gerson, 9:26-34; *see generally id.*, 9:24-10:10 (describing echo cancellation).)

234. A person of ordinary skill in the art would have understood that Gerson discloses “**adaptive echo cancellation**.” To start, I note that, like DeLine, the Examiner cited Gerson during prosecution of the ’720 patent for its disclosures related to this limitation, which the applicant did not dispute. (EX1019, pp.00357-00359 (pending claims 19 and 39 reciting “adaptive echo cancellation”), 00413 (office action rejecting claims 19 and 39 in view of Gerson), 00441-00442 (applicant response).) The Examiner observed that “Gerson, in an analogous field of speech-controlled systems, teaches an echo canceling block capable of eliminating echoes and noise in varying conditions that provides the added benefit to the other prior art of preventing an echo from being recognized at a speech recognizer (*Col. 9, Line 24-Col.10, Line 10*).” (EX1019, p.00413.) A person of ordinary skill in the art would have understood Gerson’s echo cancellation to be “**adaptive**” because, as the Examiner noted, it eliminated echoes and noise in varying conditions.

235. Moreover, a person of ordinary skill in the art would have understood that Gerson's echo cancellation technique—"acoustic echo cancellation"—is "**adaptive**." For example, *Digital Signal Processing in Telecommunications* (1996) [EX1012], explains that "[a]coustic echo cancellation works by forming in an adaptive manner a replica of the echo path." (EX1012, p.51 (section 3.2.2); *see also id.*, ("the filter has to be adaptive").) Similarly, U.S. Patent No. 5,561,668, which as noted, Gerson cites (and incorporates by reference) as an example of "acoustic echo cancellation," states that it involves adaptation. (EX1013, 4:7-10 ("The convolvers 29a-p are preferably configured for adapting to changes in the respective subbands of the echo path by periodically updating the stored impulse response coefficients."), Figs. 3 & 4).)

236. **Rationale and Motivation to Combine (Ground 2 Prior Art and Gerson)**. It would have been obvious to a person of ordinary skill in the art to combine Gerson with Coffman and its conversational computing system under the proposed combination, predictably resulting in the speech unit of a PVA in Coffman's conversational system as taught by the combination of Coffman and DeLine further using "**adaptive echo cancellation**." Combining in this manner would have therefore resulted in a speech unit that included a filter that "**employs adaptive echo cancellation to optimize the signal to noise ratio of the electronic**

signal” as claimed.

237. Gerson is an analogous reference in the same field of speech recognition and processing as the '720 patent. ('720, 1:12-21; Gerson, e.g., 1:8-13.) Gerson would have also been reasonably pertinent to a number of problems facing the patent inventors, including how to improve the quality of speech signals received in a vehicle environment or other environment with ambient noise, including from the system itself, such as synthesized speech.

238. A person of ordinary skill in the art would have been motivated to further combine with Gerson. As noted, Gerson explains that it is “human nature” to barge-in during presentation of an output audio signal. (Gerson, 1:35-39.) Gerson’s use of acoustic echo cancellation, by allowing barge-in, advantageously enables a user to have a more natural and efficient interaction with a speech-based system. Because Coffman likewise aims to provide a conversational computing system in which a user and the system interact using speech/audio, a person of ordinary skill in the art would have readily appreciated the applicability of Gerson’s teachings and the benefits that would be achieved by incorporating them. Moreover, with respect to the speech unit in Coffman’s conversational car PC (PVA) in particular, Gerson notes that its techniques can be applied “with particular advantage” to in-vehicle systems (Gerson, 6:10-12), further confirming and

motivating the combination. A person of ordinary skill in the art therefore would have been motivated to incorporate acoustic echo cancellation into the similar environment contemplated by Coffman to obtain the same benefits and advantages (or found it at least obvious to try in Coffman's system).

239. A person of ordinary skill in the art would have had a reasonable expectation of success. The incorporation of acoustic echo cancellation as taught by Gerson would have involved the application of conventional techniques and technologies. As noted, Gerson describes for example that acoustic echo cancellation "is well known in the art." (Gerson, 9:26-34.) A person of ordinary skill in the art would have found the disclosures of Coffman, DeLine and Gerson sufficiently detailed to implement the techniques on which I rely, without undue experimentation.

7. **Claim 34: "The method according to 32, wherein the filter employs at least one of adaptive echo cancellation or adaptive lossy audio compression to optimize the signal to noise ratio of the electronic signal."**

240. For purposes of my analysis, claim 34 is not meaningfully different from claim 19, and it therefore would have been obvious for the same reasons.

VI. NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

241. As explained in **Part III.B**, I understand that "secondary considerations" may be used as indicia of nonobviousness, provided that any such

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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 claims addressed in this Declaration are not obvious. To the extent 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

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

243. I hereby declare that all statements made herein of my own knowledge

Declaration of Christopher M. Schmandt in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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.
