

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

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 7,693,720 B2**

Table of Contents

	Page
I. MANDATORY NOTICES UNDER §42.8(A)(1)	17
A. Real Party-In-Interest under §42.8(b)(1).....	17
B. Related Matters under §42.8(b)(2)	17
C. Lead and Back-Up Counsel under §42.8(b)(3)	18
D. Service Information	19
II. FEE PAYMENT	19
III. REQUIREMENTS UNDER §§ 42.104 AND 42.108 AND CONSIDERATIONS UNDER §§ 314(A) AND 325(D).....	19
A. Grounds for Standing	19
B. Identification of Challenge and Statement of Precise Relief Requested	19
C. Considerations Under §§ 314(a) and 325(d).....	20
IV. OVERVIEW OF THE CHALLENGED PATENT.....	21
A. Person of Ordinary Skill in the Art (POSITA)	21
B. Specification Overview	21
V. CLAIM CONSTRUCTION	23
VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE.....	23
A. Overview of Grounds	23
B. Ground 1: Claims 1-16, 22-25, 27, 29-31, 35-43, 52, 55 Are Obvious Over Coffman in View of Kanevsky, Kneser, and Zadrozny.....	24
1. Claim 1: “A mobile system...” (1[pre])	24
(a) “a speech unit...” (1[a]).....	25
(b) “a natural language speech processing system...” (1[b])	28
(i) “a speech recognition engine...” (1[b][i]).....	35
(ii) “a parser...” (1[b][ii]).....	42

Table of Contents (continued)

	Page
(1) “determining...” (1[b][ii][1]).....	43
(2) “selecting...” (1[b][ii][2]).....	43
(3) “transforming...” (1[b][ii][3])	44
(iii) “an agent architecture...” (1[b][iii]).....	48
2. Claim 2 (2[pre])	55
(a) “sends and receives events...” (2[a])	56
(b) “provides a multi-threaded environment...” (2[b]).....	58
3. Claim 3	61
4. Claim 4.....	61
5. Claim 5.....	62
6. Claim 6.....	63
7. Claim 7.....	65
8. Claim 8.....	65
9. Claim 9.....	66
10. Claim 10.....	68
11. Claim 11	69
12. Claim 12	70
13. Claim 13.....	70
14. Claim 14.....	71
15. Claim 15	71
16. Claim 16.....	72
17. Claim 22.....	73
18. Claim 23	73
19. Claim 24.....	74
20. Claim 25.....	74
21. Claim 27	75
22. Claim 29 (29[pre])	77

Table of Contents (continued)

	Page
(a) “determines...” (29[a])	77
(b) “tags...” (29[b])	77
23. Claim 30	78
24. Independent Claim 31: “A method...” (31[pre])	79
(a) “receiving...” (31[a])	79
(b) “recognizing...” (31[b])	79
(c) “determining...” (31[c])	79
(d) “selecting...” (31[d])	80
(e) “transforming...” (31[e])	80
(f) “forwarding...” (31[f])	80
25. Claim 35	81
26. Claims 36-37	82
27. Claim 38	82
28. Claim 39	82
29. Claim 40	83
30. Claim 41	83
31. Claim 42	83
32. Claim 43	84
33. Claims 52, 55	84
C. Ground 2: Claims 18, 20-21, 32-33 Are Obvious Over Ground 1 Prior Art in Further View of DeLine	85
1. Claim 18	85
2. Claim 20	91
3. Claim 21	91
4. Claim 32	92
5. Claim 33	92

Table of Contents
(continued)

	Page
D. Ground 3: Claims 19 and 34 Are Obvious Over Ground 2 Prior Art in Further View of Gerson.	93
1. Claim 19	93
2. Claim 34	96
VII. CONCLUSION.....	96
CERTIFICATE OF SERVICE	98

Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

List of Exhibits

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 ”)
1002	Declaration of Christopher M. Schmandt (“ Schmandt ”)
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 ”)
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 ”)
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 ”)

Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

List of Exhibits

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)
1018	Declaration of Ingrid Hsieh-Yee, Ph.D.
1019	Prosecution history of U.S. Patent No. 7,693,720 B2
1020	Proof of Service of Complaint, <i>Dialect, LLC v. Google LLC</i> , Case No. 1:23-cv-00378 (D. Del.)
1021	International Preliminary Examination Report for Maes (EX1007) application

Claim Listing

	Claim 1
1[pre]	1. A mobile system responsive to a user generated natural language speech utterance, comprising:
1[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[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][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][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][1]	determining a context for the natural language speech utterance;
1[b][ii][2]	selecting at least one of the plurality of domain agents based on the determined context; and
1[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
1[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

Claim Listing

	create a response to the formulated question or command and format the response for presentation to the user.
	Claim 2
2[pre]	2. The mobile system according to claim 1, wherein the natural language speech processing system further includes an event manager that:
2[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[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.
	Claim 3
	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.
	Claim 4
	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.
	Claim 5
	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.

Claim Listing

	Claim 6
	6. The mobile system according to claim 1, wherein the selected domain agent includes data for communicating with one or more devices.
	Claim 7
	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.
	Claim 8
	8. The mobile system according to claim 6, wherein at least one of the one or more devices is connected to the vehicle.
	Claim 9
	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.
	Claim 10
	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.
	Claim 11
	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.
	Claim 12
	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.

Claim Listing

	Claim 13
	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.
	Claim 14
	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.
	Claim 15
	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.
	Claim 16
	16. The mobile system according to claim 15, wherein the communication network is a wide area wireless network.
	Claim 18
	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.
	Claim 19
	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.
	Claim 20
	20. The mobile system according to claim 18, wherein the array microphone is at least a one-dimensional array.

Claim Listing

	Claim 21
	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.
	Claim 22
	22. The mobile system according to claim 1, wherein the speech unit is located remotely from the natural language speech processing system.
	Claim 23
	23. The mobile system according to claim 1, further comprising at least one of a display or a keypad.
	Claim 24
	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.
	Claim 25
	25. The mobile system according to claim 1, wherein the communicatively coupled services include at least one shared network resource.
	Claim 27
	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.
	Claim 29
29[pre]	29. The mobile system according to claim 1, wherein the speech recognition engine:
29[a]	determines an identity of the user based on unique voice characteristics of the user; and

Claim Listing

29[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.
	Claim 30
	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.
	Claim 31
31[pre]	31. A method responsive to a user generated natural language speech utterance, comprising:
31[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[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[c]	determining, at a parser connected to the computer device on the vehicle, a context for the natural language speech utterance;
31[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[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

Claim Listing

31[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.
	Claim 32
	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.
	Claim 33
	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.
	Claim 34
	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.
	Claim 35
	35. The method according to claim 31, the wherein the selected domain, 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.
	Claim 36
	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.

Claim Listing

	Claim 37
	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.
	Claim 38
	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.
	Claim 39
	39. The method according to claim 38, wherein the remotely located service is associated with a remotely located device.
	Claim 40
	40. The method according to claim 38, wherein the request is transmitted to the remotely located service via a communication network.
	Claim 41
	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.
	Claim 42
	42. The method according to 38, wherein the request is transmitted to the remotely located service via a wide-area RF transceiver.
	Claim 43
	43. The method according to claim 31, wherein the communicatively coupled services include one or more shared network resources.
	Claim 52
	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.

Claim Listing

	Claim 55
	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.

I. MANDATORY NOTICES UNDER §42.8(A)(1)

A. Real Party-In-Interest under §42.8(b)(1)

Google LLC (“Petitioner”) is the real party-in-interest to this IPR petition.¹

B. Related Matters under §42.8(b)(2)

The ’720 patent is the subject of pending litigation in which Petitioner is named as a defendant, *Dialect, LLC v. Google LLC*, Case No. 1:23-cv-00378 (D. Del.) (filed April 3, 2023). Petitioner was first served with a complaint alleging infringement of the ’720 patent on April 6, 2023. (EX1020, p.002.) On March 28, 2024, Petitioner’s motion to transfer to the Northern District of California was granted. *See* Dkt. 32, Case No. 1:23-cv-00378 (D. Del.). The ’720 patent is also the subject of *Dialect, LLC v. Amazon.com, Inc.*, Case No. 1:23-cv-00581-LMB-LRV (E.D. Va.) (filed May 1, 2023).

Petitioner is filing IPR petitions against eight patents asserted by Patent Owner in the pending litigation involving Petitioner:

U.S. Patent No.	IPR Case
7,398,209 B2	IPR2024-00746
7,502,738 B2	IPR2024-00747
8,015,006 B2	IPR2024-00748

¹ Google LLC is a subsidiary of XXVI Holdings Inc., which is a subsidiary of Alphabet Inc. XXVI Holdings Inc. and Alphabet Inc. are not real parties in interest to this proceeding.

Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

U.S. Patent No.	IPR Case
7,693,720 B2	IPR2024-00749
9,031,845 B2	IPR2024-00750
8,447,607 B2	IPR2024-00751
8,849,652 B2	IPR2024-00752
7,640,160 B2	IPR2024-00753

C. Lead and Back-Up Counsel under §42.8(b)(3)

Petitioner provides the following designation of counsel.

LEAD COUNSEL	BACK-UP COUNSEL
Heidi L. Keefe (Reg. No. 40,673) hkeefe@cooley.com COOLEY LLP ATTN: Patent Group 1299 Pennsylvania Ave. NW, Suite 700 Washington, DC 20004 Tel: (650) 843-5001 Fax: (650) 849-7400	Andrew C. Mace (Reg. No. 63,342) amace@cooley.com COOLEY LLP ATTN: Patent Group 1299 Pennsylvania Ave. NW, Suite 700 Washington D.C. 20004 Tel: (650) 843-5287 Fax: (650) 849-7400
	Mark R. Weinstein (Admission <i>pro hac vice</i> to be requested) mweinstein@cooley.com COOLEY LLP ATTN: Patent Group 1299 Pennsylvania Avenue NW Suite 700 Washington D.C. 20004 Tel: (650) 843-5007 Fax: (650) 849-7400
	Reuben Chen (Admission <i>pro hac vice</i> to be requested) rchen@cooley.com COOLEY LLP

LEAD COUNSEL	BACK-UP COUNSEL
	ATTN: Patent Group 1299 Pennsylvania Avenue NW Suite 700 Washington D.C. 20004 Tel: (650) 843-5007 Fax: (650) 849-7400

D. Service Information

This Petition is being served by Federal Express to the correspondence address of record for the '720 patent, 174899 - The Law Offices of David A. Gerasimow, P.C., 211 W. Wacker Dr., Ste. 1717, Chicago, IL 60606. Petitioner consents to electronic service at the addresses provided above for lead and back-up counsel.

II. FEE PAYMENT

Petitioner requests review of 42 claims, with a \$66,250 payment.

III. REQUIREMENTS UNDER §§ 42.104 AND 42.108 AND CONSIDERATIONS UNDER §§ 314(A) AND 325(D)

A. Grounds for Standing

Petitioner certifies that the '720 patent is available for IPR and that Petitioner is not barred or otherwise estopped.

B. Identification of Challenge and Statement of Precise Relief Requested

Petitioner requests IPR institution based on:

Petition for *Inter Partes* Review of
U.S. Patent No. 7,693,720 B2

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), Zadrozny (EX1006)
2	18, 20-21, 32-33	Ground 1 Prior Art + DeLine (EX1010)
3	19, 34	Ground 2 Prior Art + Gerson (EX1011)

C. Considerations Under §§ 314(a) and 325(d)

§314(a): Petitioner is unaware of another post-grant challenge against the '720 patent. The pending litigation against Petitioner is still in early stages, with no discovery or claim construction having occurred and no trial date set, and as noted, was recently transferred to the Northern District of California. If IPR is instituted, Petitioner intends to bring a motion to stay the litigation.

§325(d): During prosecution of the '720 patent, an applicant-submitted IDS identified Coffman's U.S. counterpart (i.e., 7,137,126). (EX1019, pp.00250-00251.) It was not discussed. Why the Examiner did not appreciate its significance is unclear, but the applicant had already disclosed nearly one hundred other references and the Examiner had already issued an office action rejecting pending claims over different prior art before it was disclosed. (EX1019, pp.00118-00123, 00205-221.) The Examiner also did not have the Kanevsky, Kneser and Zadrozny references that this Petition relies on, which provide further context and detail for understanding Coffman and how its disclosures provide compelling grounds of obviousness.

DeLine and Gerson were cited for disclosing limitations in certain dependent claims, which the applicant did not dispute. (EX1019, pp.00357-00359 (pending claims), 00412-00413 (rejections based on DeLine and Gerson), 00441-00442 (applicant response).) The Petition cites DeLine and Gerson against those dependent claims here.

IV. OVERVIEW OF THE CHALLENGED PATENT

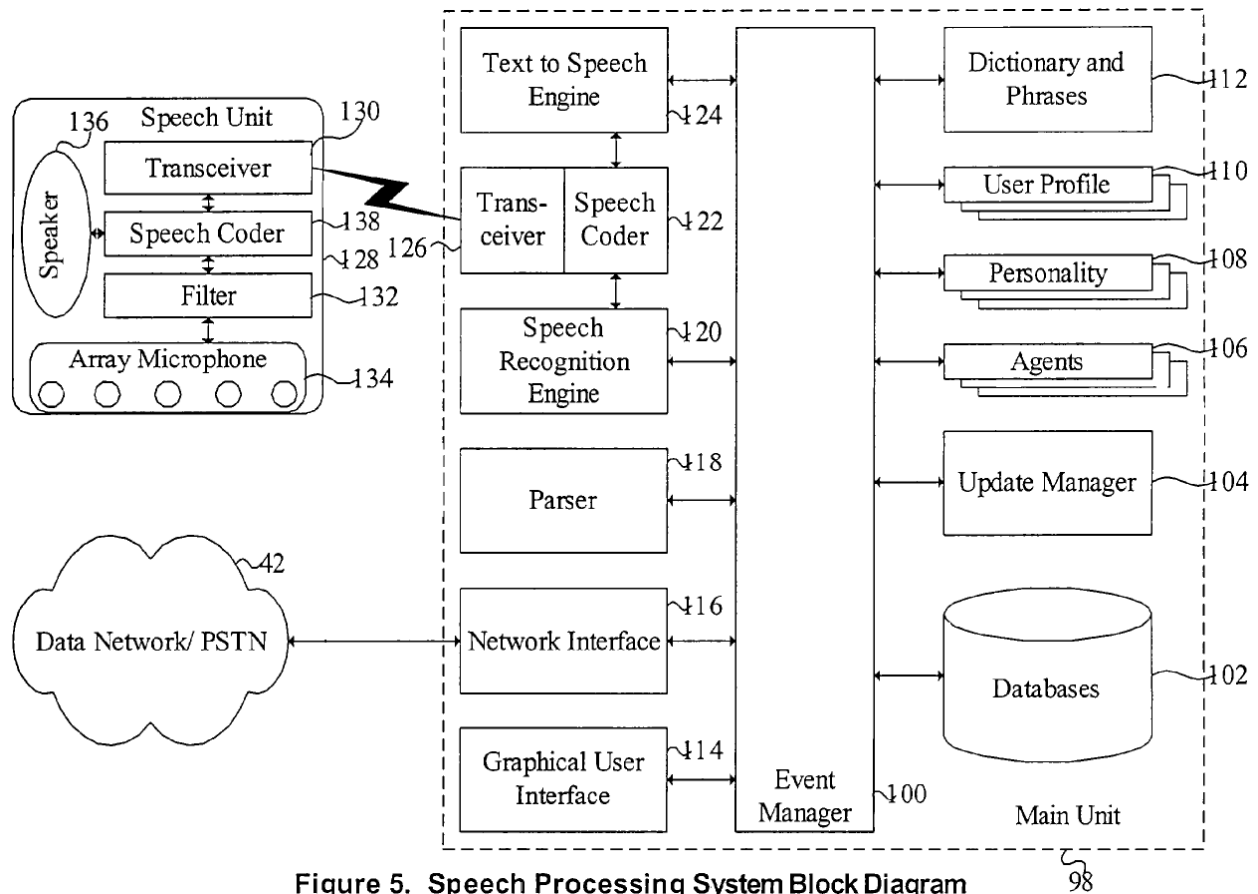
A. Person of Ordinary Skill in the Art (POSITA)

A POSITA as of July 15, 2002² 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 (and more experience could substitute for less formal education, and vice versa). (EX1002 ("Schmandt"), ¶¶11-15.)

B. Specification Overview

Figure 5 of the '720 patent, below, "is an overall diagrammatic view of [the] interactive natural language speech processing system according to an embodiment of the invention":

² Petitioner does not concede a July 15, 2002 priority date, but that issue is irrelevant here because the prior art predates that date.



(’720, 15:44-46, Fig. 5.) As shown, the system includes 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 user’s spoken utterance. (’720, 20:28-30.) Speech recognition engine **120** processes the speech “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.) To select the agent to invoke, “[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, 27:60-66, 28:25-27.)

V. CLAIM CONSTRUCTION

Express claim construction is unnecessary because the prior art renders the challenged claims obvious under any reasonable construction. Petitioner reserves its right to respond to Patent Owner claim construction arguments.

VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE

A. Overview of Grounds

The challenged claims attempt to lay claim to well-known techniques for speech-enabled interfaces for interacting with and directing computer systems. The claims employ standard techniques known before the ’720 patent, such as using speech recognition and natural language processing to select an application for carrying out a query and/or command included in a spoken utterance.

All grounds rely primarily on **Coffman** (**EX1003**), which discloses a remarkably similar conversational computing system. All grounds combine Coffman with **Kanevsky** (**EX1004**), **Kneser** (**EX1005**) and **Zadrozny** (**EX1006**), which disclose known techniques related to speech recognition and natural language processing. As detailed below, Coffman, Kanevsky, and Zadrozny share links and thus comprise a collective body of work describing aspects of the same systems and

methods—for example, Coffman expressly incorporates Kanevsky by reference (Coffman, 37:10-13), and Kanevsky in turn discusses Zadrozny (Kanevsky, 14:22-25); Coffman and Kanevsky share a named co-inventor (Stephane Maes); and the named inventors and authors of Coffman, Kanevsky, and Zadrozny are identified as associated with IBM. *See Paice, LLC v. Ford Motor Co.*, 881 F.3d 894, 906-907 (Fed. Cir. 2018) (incorporated material “effectively part of the host document as if it were explicitly contained therein”); *Boston Scientific Scimed, Inc. v. Cordis Corp.*, 554 F.3d 982, 991 (Fed. Cir. 2009) (“[c]ombining two embodiments disclosed adjacent to each other in a prior art patent does not require a leap of inventiveness”). Grounds 2-3 combine these references with **DeLine** (EX1010) and **Gerson** (EX1011) for features recited in dependent claims 18-21 and 32-34.

The references are prior art under at least pre-AIA §§102(b) (Coffman, Kanevsky, Kneser, Zadrozny) and 102(e) (DeLine, Gerson). (*See* EX1018, ¶¶25-48 (Zadrozny).)

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

1. Claim 1: “A mobile system...” (1[pre])

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.)³ As shown below, the claimed system is disclosed and rendered obvious by the prior art.

(a) “a speech unit...” (1[a])

The ’720 patent describes a “**speech unit**” as comprising a microphone and speaker, with optional components. (’720, 20:28-35, 20:35-39, 20:52-55, Fig. 5; Schmandt, ¶75.) The speech unit and a main unit processing device can be integrated into a single physical device or connected via wires or wirelessly. (’720, 18:62-64; *see also id.*, Fig. 3, 18:53-57, 3:1-5, 20:33-35, 20:52-55, Fig. 5; Schmandt, ¶75.)

A POSITA would have understood and found it obvious that Coffman’s system would comprise “**a speech unit connected to a computer device.**” (Schmandt, ¶76.) 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 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)”), 37:14-22 (“Spoken input from a speech client”), 35:2-3 (“listen to the output of a task”), 39:12 (“audio only output”), 39:30 (“voice output”), Fig. 6

³ All underlining added to quoted material unless noted.

(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, 12:5-7, 13:1-3, 56:21-25, 28:29-31, Figs. 1 & 6.)

A POSITA 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.**” (Schmandt, ¶77.) As noted, the computer device that receives speech input can take the form of a “conversational car PC (PVA, Personal Vehicle Assistant).” (Coffman, 56:21-25, 55:27-28.)

Coffman further discloses “**the speech unit receives a...speech utterance from a user.**” (Schmandt, ¶78.) Coffman describes receiving spoken input from a person using, for example, a telephone or microphone, illustrated below:

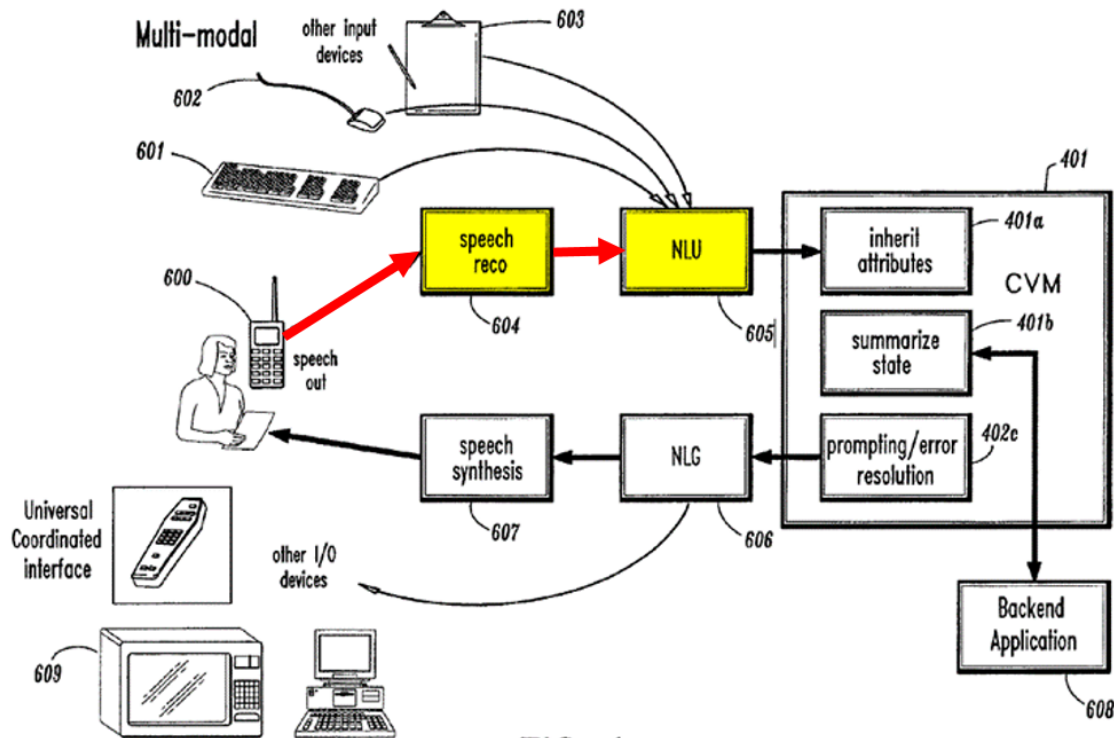
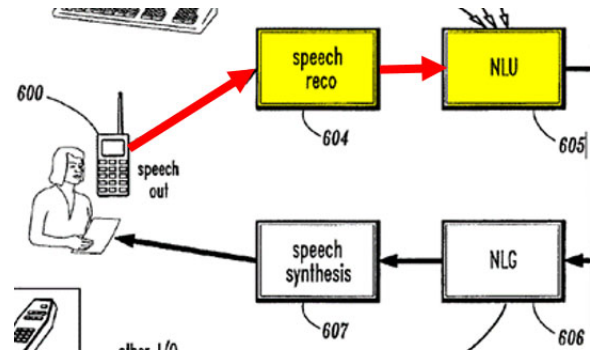


FIG. 6

(Coffman, Fig. 6 (annotations added); *see also id.*, 37:14-22 (“Spoken input from a speech client (e.g., telephone 600)...”), 53:8-10, 41:2-5, 53:22-26.)

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 refers to the forms of language a person uses for communication (e.g., English). (Schmandt, ¶79 (citing EX1008, p.0035 (“natural language” entry)).) Coffman discloses several examples of natural language speech utterances. (Coffman, 6:32-7:3 (“the command ‘I would like to open this’”), 48:1-3 (“e-mail from my mother”), 56:25-57:2 (“grocery list,” “it should be synchronized with my PDA,” and “Do it!”); Schmandt, ¶79.)

Coffman's receipt (and use) of spoken "**natural language**" input is confirmed by Coffman's explanation that spoken input, once recognized, is subjected to natural language parsing and understanding ("NLU") processes. (Coffman, e.g., Fig. 6 (annotated excerpt at right), 53:22-26 ("parsing and NLU"), 53:8-10, 37:14-22, 6:6-8, 37:29-32, 5:1-4; Schmandt, ¶80.)

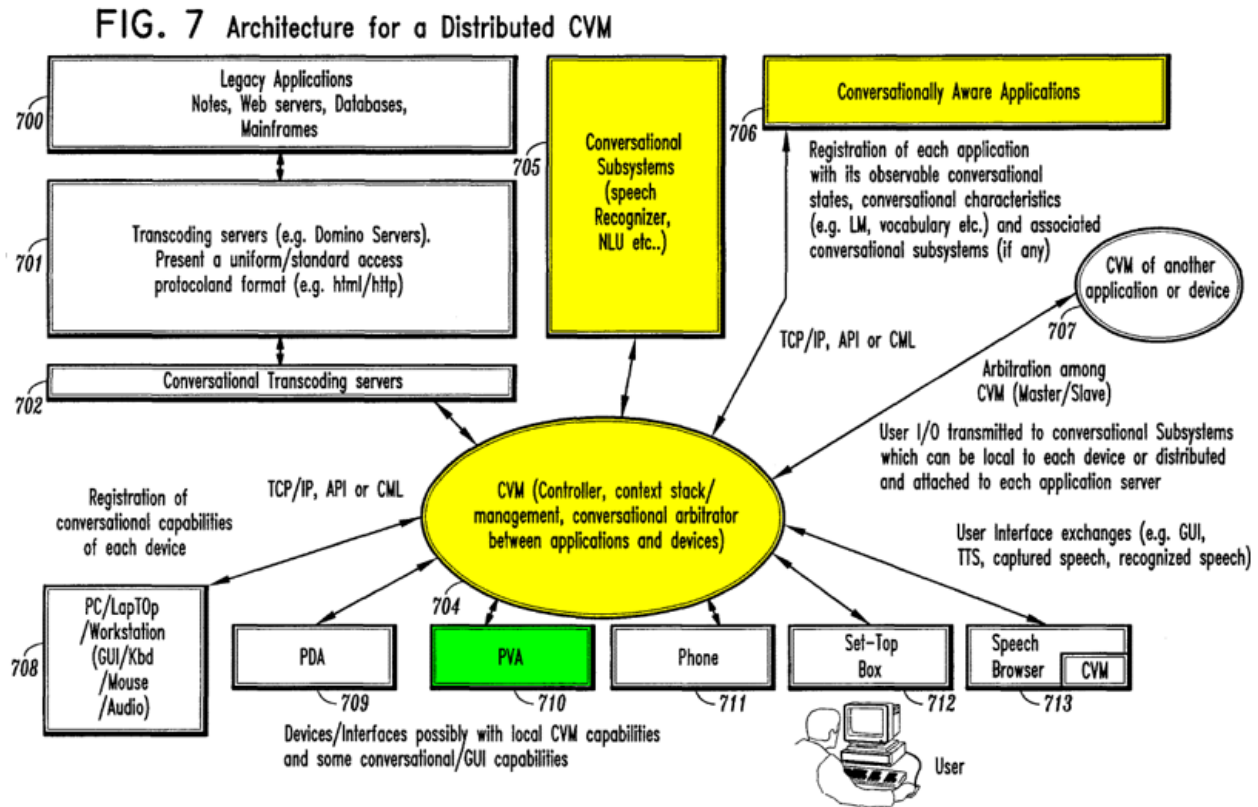


A POSITA would further have understood and found it obvious for Coffman's "**speech unit**," as part of the functionality for processing incoming speech, to "**convert[] the received natural language speech utterance into an electronic signal.**" (Schmandt, ¶81.) In a speech-processing system using conventional digital computing components, such as Coffman's, a POSITA would have understood that, to process a user's utterance, the utterance would need to be converted into an electronic speech signal. (Schmandt, ¶81 (citing Coffman, 13:1-3, 26:11-13; EX1009, pp.005 ("analog-to-digital converter"), 0013 ("digital signal")).)

(b) "a natural language speech processing system..." (1[b])

In Coffman, the claimed "**natural language speech processing system**" connected to the 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 communicate with each other. (Schmandt, ¶83.) Coffman Figure 7 illustrates these components and their connection to the PVA:



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

Coffman Figure 4 further illustrates how Coffman’s “conversational subsystems,” in turn, comprise a speech recognition engine, natural language understanding and natural language parsing system, and a text-to-speech synthesis (TTS) system:

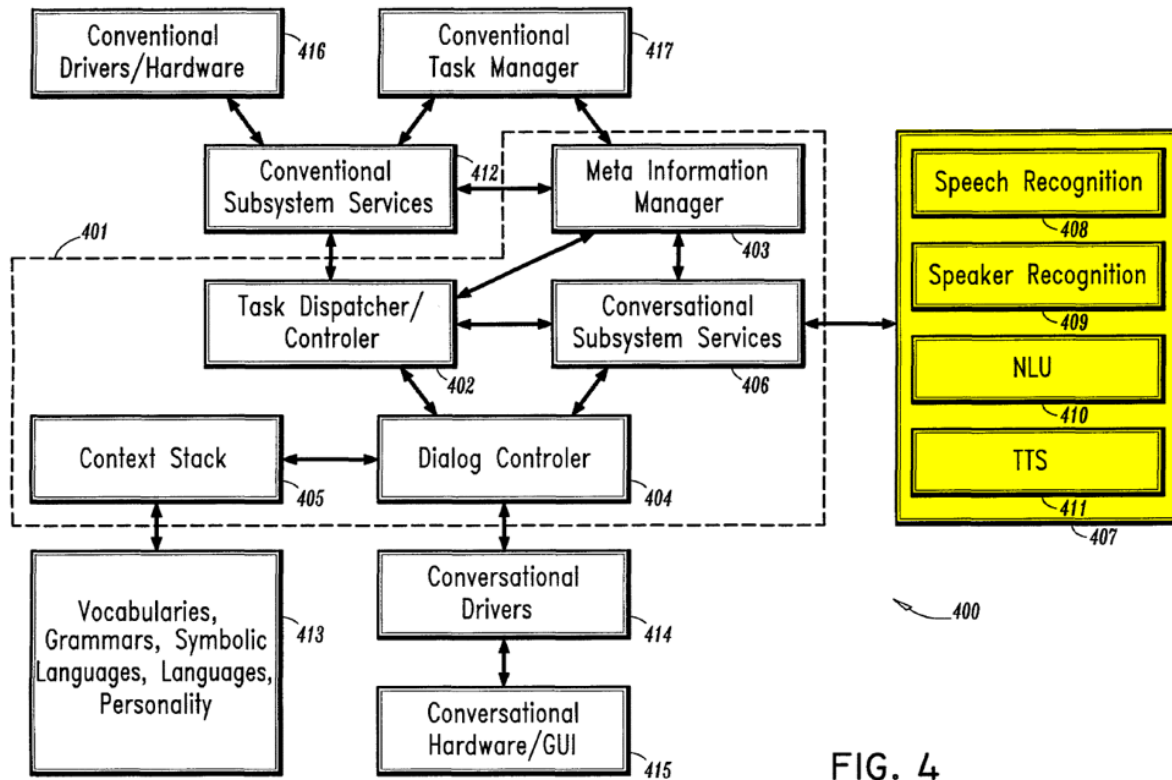
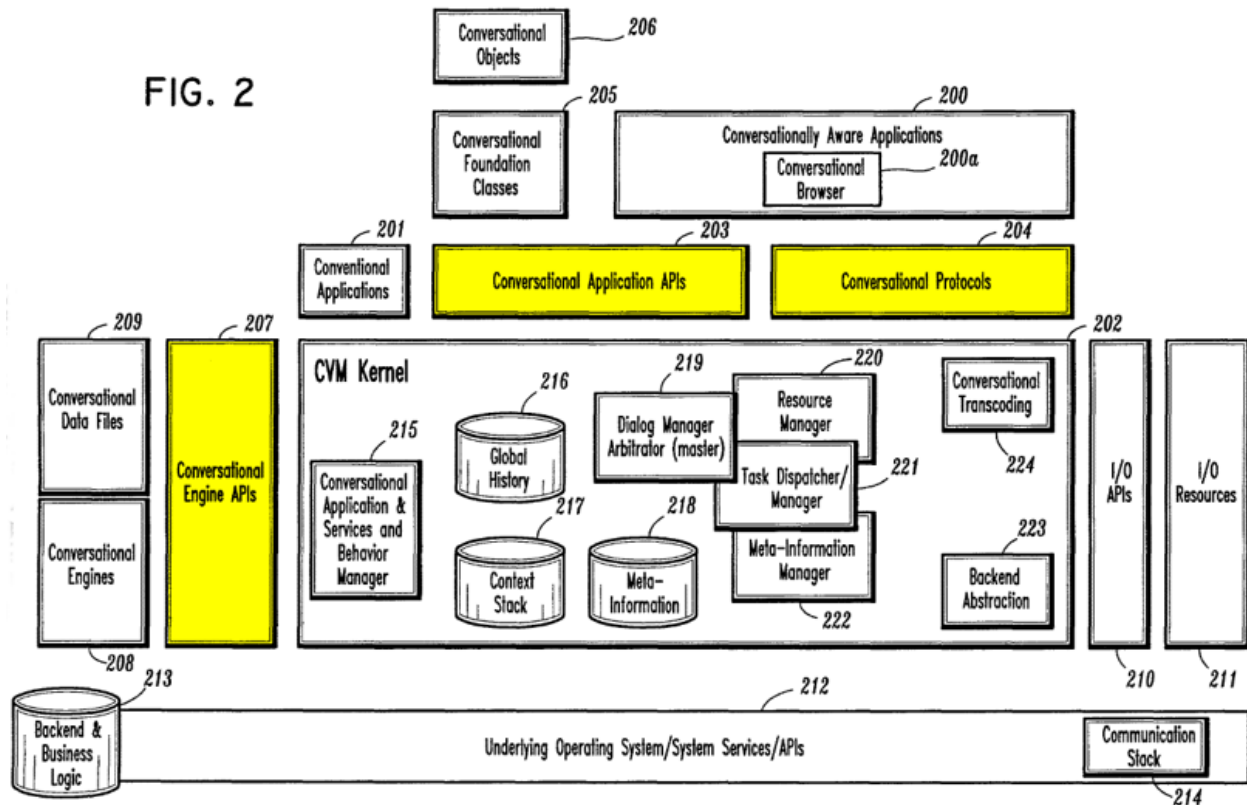


FIG. 4

(Coffman, Fig. 4 (highlighting added); *see also id.*, 29:1-4.) Coffman’s speech recognition engine and natural language understanding and natural language parsing system are addressed in the analysis of 1[b][i] and 1[b][ii], which recite “**a speech recognition engine**” and “**a parser**,” respectively.

Regarding the system’s conversational APIs and protocols, Coffman explains that “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.) Coffman Figure 2 illustrates:



(Coffman, Fig. 2 (highlighting added).) As shown, the APIs include both “conversational application APIs,” which enable a conversationally-aware application to communicate with a CVM, and “conversational engine APIs,” which enable communication with conversational subsystems. (Coffman, 5:1-6, 19:16-19, 12:28-30.) 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.)

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 PVA (“**the computer device on the vehicle**”). (Coffman, 13:1-3, 29:1-4; Schmandt, ¶86.) For example, as noted, Coffman Figure 7 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 POSITA 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. (Schmandt, ¶86.)

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. (Schmandt, ¶87.) 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”).) The analysis of 1[b][i]-[iii] discusses 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. (Coffman, 13:8-11, Fig. 1; Schmandt, ¶87.)

Turning to why a conversationally-aware application is a “**domain agent**,” in the software context, “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. (Schmandt, ¶88.) The ’720 patent describes “agents” as “executables that receive, process and respond to user questions, queries and commands” and states that they “provide convenient and re-distributable packages or modules of functionality, typically for a specific domain.” (’720, 4:7-24.)

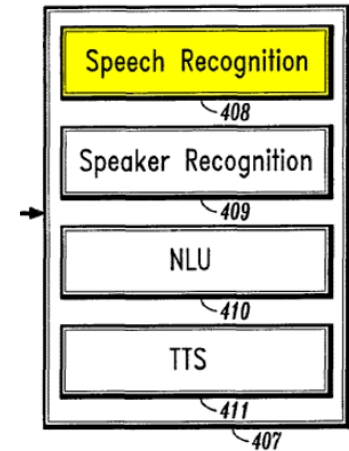
A conversationally-aware application in Coffman—the executable software and associated data, such as vocabulary, grammar, language models and NLU symbolic mapping—is a “**domain agent**.” (Coffman, e.g., 13:26-29; Schmandt, ¶89.) A conversationally-aware application receives, processes, and responds 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”); Schmandt, ¶89.) It 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, 16:35-17:6 (listing “application specific” foundation classes), 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 (“...capability to have a natural dialog with NLU, NLG and mixed initiative across multiple applications, with multiple domains.”); Schmandt, ¶¶89-90.) Coffman also describes development tools for building complete and re-distributable applications. (Coffman, 58:12-20; *see also id.*, 58:27-28, 7:29-8:9, 52:9-11 & 52:18-21 (describing agents to upgrade or install applications).)

As explained next, 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...” (1[b][i])**

Coffman’s system includes various “conversational subsystems,” including a **“speech recognition engine.”** (Coffman, 10:28-30, 29:1-4 (“conversational resource subsystems (engines) 407 including, for example, a speech recognition system 408”), Fig. 4 (annotated and excerpted at right).) The speech recognition engine **“recognizes at**



least one of words or phrases from the electronic signal.” (Coffman, 37:20-22 (“Spoken input from a speech client (e.g., telephone 600) is subject to a speech recognition process 604...”); *see also id.*, Fig. 6 (depicting input to and output from “speech reco” 604), 38:9-11; Schmandt, ¶94.)

A POSITA would have understood and found it obvious that 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.”** (Schmandt, ¶95.) As explained below, this data in Coffman comprises n-gram language model data used for speech recognition that is provided by conversationally-aware applications.

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.) For background, in a speech recognition context, a “language model” refers to a formal representation of the structure of a natural language. (Schmandt, ¶96 (citing EX1008, p.0031 (“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.” (*Id.* (EX1008, quoting p.0031, citing pp.0035 (“n-gram language model”), 0050 (“stochastic language model”)).) A “trigram” language model, for example, therefore refers to an n-gram language model that uses the probabilities of the occurrence of sequences of three words. (Schmandt, ¶96 (citing EX1008, p.0053 (“trigram”)).)

Because using language models for speech recognition was well known, Coffman unsurprisingly does not offer much specific detail. (Schmandt, ¶97.) *See Bayer Healthcare LLC v. Baxalta Inc.*, 989 F. 3d 964, 982 (Fed. Cir. 2021) (“[a] patent need not teach, and preferably omits, what is well known in the art”). Coffman does incorporate **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; *see also id.*, 13:48-61.)

Under the 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). (Schmandt, ¶98 (citing EX1008, p.0053 (“trigram”)).) The trigram data in a language model comprises “**a plurality of dictionary and phrase entries**” because a POSITA would have understood and found it obvious to organize the n-gram language model(s) used by the Coffman’s system as a table, with the trigram data being “**entries**” in the table. (Schmandt, ¶98 (citing EX1016, p.006 (with n-gram language model, “syntactic knowledge is captured in a large table of word bigrams (pairs), trigrams (triples),...N-grams”), EX1014, ¶0040 (“N-gram language model [] works from tables of probabilities of sequences of words.”)).) The table contains an entry for a particular word, with a link to other word(s) to form an n-gram. (*Id.*) 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 (“**dictionary**”) and sequences of words (“**phrase(s)**”) to refer to during speech recognition. (Schmandt, ¶98 (citing EX1015, p.005 (“dictionary” entry)).)

Rationale and Motivation to Combine (Coffman and Kanevsky). It would have been obvious to combine Coffman and Kanevsky, predictably resulting in Coffman’s system in which language models provided by conversationally-aware applications and used for speech recognition were n-gram language models (e.g.,

trigram language model). (Schmandt, ¶99) Combining the teachings of Coffman and Kanevsky in this manner would have 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. (*Id.*)

Both Coffman and Kanevsky are analogous references in the field of the ’720 patent and would have been reasonably pertinent to problems facing the patent inventors. (Schmandt, ¶100; ’720, 1:12-21; Coffman, 1:7-8, Figs. 1, 4 & 6; Kanevsky, 1:15-21.)

A POSITA would have been motivated to combine. (Schmandt, ¶101.) First, Coffman expressly incorporates Kanevsky (Coffman, 37:10-13) and is therefore part of Coffman’s disclosures for purposes of applying it as prior art. *See, e.g., Paice*, 881 F.3d at 906-907; *see also Boston Scientific Scimed*, 554 F.3d at 991. Moreover, both references share a named co-inventor, Stephane Maes, and identify IBM as the assignee/applicant. (Coffman, Kanevsky face pages.) Therefore, a POSITA would have consulted both of these references to understand the features and functionality of Coffman’s system and/or actually implement or adapt the system and features it provides. (Schmandt, ¶101.)

Moreover, Coffman already discloses that a conversationally-aware application provides its “language models” to the system in connection with

processing a user's input. A POSITA would have understood that a language model in this context likely refers to an n-gram language model, and Kanevsky provides confirmation, explaining that a speech recognition engine can be implemented using a trigram language model. (Schmandt, ¶102 (citing EX1008, pp.0031 ("language model"), 0035 ("n-gram language model"), Kanevsky, 13:50-56).) Moreover, the use of an n-gram language model as Kanevsky teaches would have involved a known technique to implement or improve Coffman's speech recognition functionality in the same way. (*Id.*)

A POSITA would have had a reasonable expectation of success, as the combination would have involved applying well-known technologies and techniques without undue experimentation. (Schmandt, ¶103 (citing '720, 1:58-59; Kanevsky, 13:48-61).)

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 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. (Schmandt, ¶104 (citing EX1008, p.0024 ("adaptive language model")).)

Coffman itself suggests modifying language models based on user interaction, describing collection of user-specific information 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; Schmandt, ¶105.) Adaptive n-gram language models and their advantages were well known, and their use would have been obvious over **Kneser** (EX1005), which 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. (Schmandt, ¶105; Kneser, 1:22-29; *see generally id.*, 1:13-37.)

Rationale and Motivation to Combine (Coffman and Kanevsky with Kneser). It would have been obvious 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 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.**” (Schmandt, ¶106.)

Kneser is an analogous reference in the field of the '720 patent, and would have been reasonably pertinent to problems facing the patent inventors. (Schmandt, ¶107; '720, 1:12-21; Kneser, 1:5-11.)

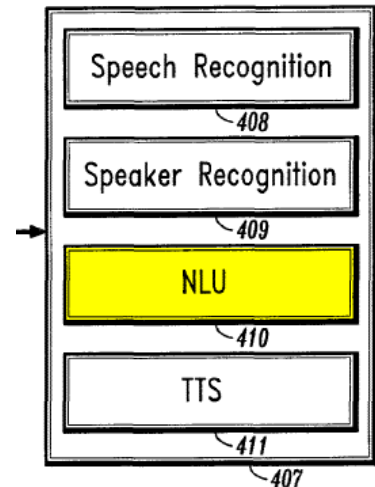
Kneser provides express motivation for incorporating “adaptive” language model techniques into Coffman’s conversational system—improving speech recognition reliability and performance. (Schmandt, ¶108.) Coffman itself confirms this motivation in the context of collecting user-specific information in order to perform “more robust recognition.” (*Id.*; Coffman, 37:5-10.) Adaptive language models as taught by Kneser provide a way in which Coffman’s system could advantageously improve speech recognition reliability and performance. Adaptive language models were well known, and a POSITA 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. (*Id.*)

A POSITA would have had a reasonable expectation of success, because the combination would have involved conventional techniques and technologies without undue experimentation. (Schmandt, ¶109.)

(ii) “a parser...” (1[b][ii])

1[b][ii] and its sub-steps (1[b][ii][1]-[3]) would have been obvious in further view of Zadrozny.

Coffman discloses “a parser.” (Schmandt, ¶111.)
Coffman’s system includes various “conversational subsystems,” including “a natural language understanding and natural language parsing system [NLU] 410.” (Coffman, 29:1-4, Fig. 4 (annotated excerpt at right).) 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 (“parsers”).)

The remainder of 1[b][ii] introduces and summarizes sub-steps 1[b][ii][1]-[3]. Therefore, the analysis of 1[b][ii][1]-[3] below explains 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, 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.

(Schmandt, ¶112; 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...**” (1[b][ii][1])

A POSITA 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.**” (Schmandt, ¶113.) Coffman uses NLU (i.e., parsing) as part of the process to determine a “context” that matches the user’s recognized utterance. Coffman explains that the conversational system includes a “context stack” where the “context,” including the state and history of each active process, such as an application process, is tracked. (Coffman, 29:18-30:3; *see also id.*, 30:4-10; Schmandt, ¶114.) The context stack is searched to determine a “context” that matches the user’s request based on NLU processing. (Coffman, 42:1-5, 41:2-5, 30:8-10; Schmandt, ¶114.)

(2) “**selecting...**” (1[b][ii][2])

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. (Schmandt, ¶116 (citing Coffman, 42:1-7 (“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”), 37:2-4).)

(3) “transforming...” (1[b][ii][3])

Coffman discloses “transforming the recognized words or phrases into at least one of a question or a command.” 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 (“transform queries into application calls”), 30:8-10 (“symbolic language (attribute value n-uple) which maps the actual query”), 6:8-10 (“ASCII command or query (i.e., lists of attribute-value pairs or n-uples).”); Schmandt, ¶117.) 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; Schmandt, ¶117.)

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 is 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”); Schmandt, ¶118.) A POSITA 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. (Schmandt, ¶¶118-119 (citing Coffman, 53:25-26, 29:1-4, 6:32-7:3; EX1008, pp.0035 (defining “Natural Language Processing”), 0040 (defining “parsing”)).) 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.”** (*Id.*)

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.”** (Schmandt, ¶120.) Zadrozny 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 formulate a recognized utterance as an intended action and parameters and then translate the action and parameters into “a form that can be executed by the application.” (Zadrozny, pp.0019-0020.) Zadrozny illustrates the technique with the example utterance “I want you to arrange a conference in my office at 5.” (Zadrozny, p.0021.) Using a grammar of constructions, the utterance is analyzed 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.0021-0022.) This action and parameters are then translated into a form for the application to execute (i.e., an application call/query). (Schmandt, ¶120.)

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. (Schmandt, ¶121; *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”).)

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

Zadrozny is an analogous reference and would have been reasonably pertinent to problems facing the patent inventors. (Schmandt, ¶123.)

A POSITA would have been motivated to combine. (Schmandt, ¶124.) Kanevsky cites Zadrozny as an example of a “semantic analyzer.” (Kanevsky, 14:22-25.) Zadrozny authors are identified as being associated with IBM, the assignee/applicant for Coffman and Kanevsky. (Zadrozny, p.0019 (“IBM Research Division”).) A POSITA therefore would have turned to Zadrozny for detail about how to implement the systems and techniques described in Coffman and Kanevsky. (Schmandt, ¶124.)

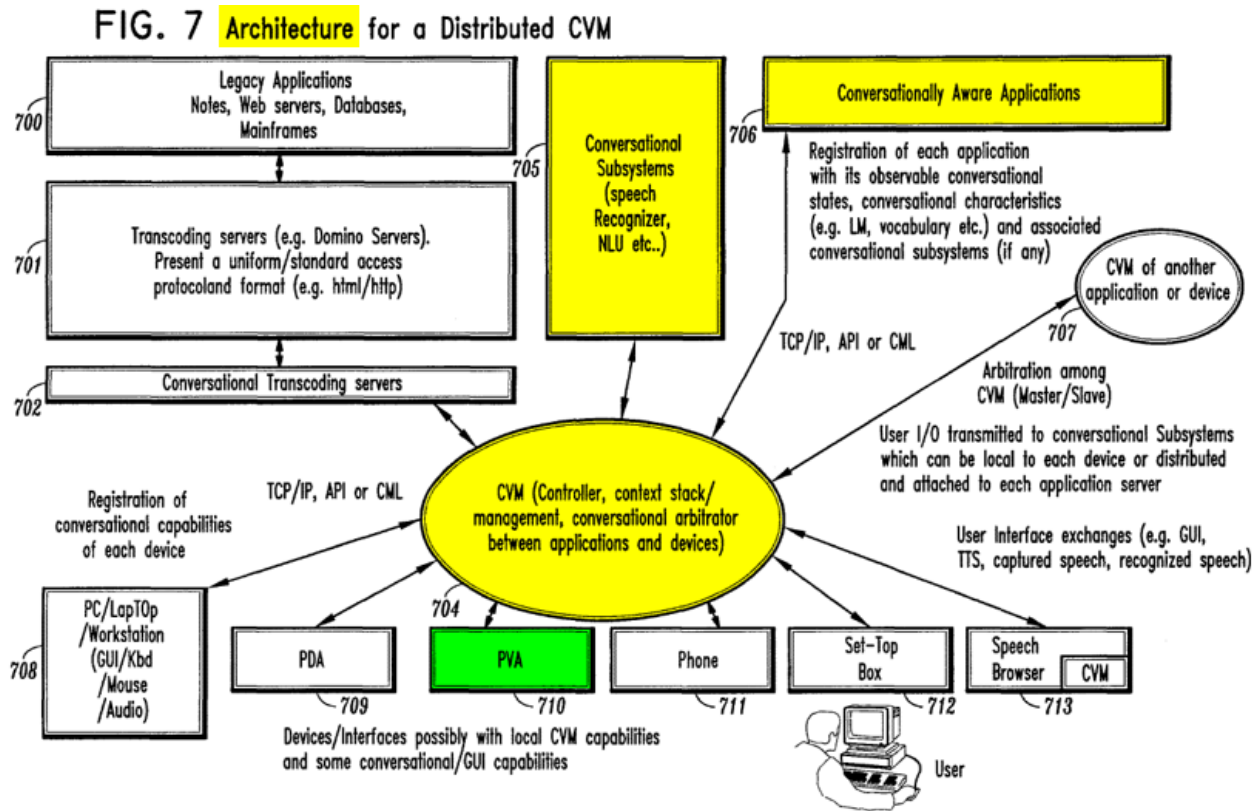
A POSITA would have appreciated that Coffman's and Zadrozny's disclosures confirm their fit. (Schmandt, ¶125.) As noted, Coffman does not describe the natural language understanding process in detail, mostly describing at a high level that applications provide grammars, vocabularies, language models 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 cites Zadrozny as an example of a semantic analyzer, a POSITA would have turned to Zadrozny for teachings regarding using grammars in natural language understanding to translate a recognized utterance into an application call. (Schmandt, ¶125.) Moreover, as noted, Zadrozny's process is just like Coffman's descriptions of formulating a user's utterance as a set of attribute-value pairs and then converting them into an application call/query. (*Id.*)

A POSITA would have had a reasonable expectation of success. (Schmandt, ¶126 (describing Zadrozny's implementation).)

(iii) "an agent architecture..." (1[b][iii])

Coffman's "**natural language speech processing system**" (CVM, conversational subsystems, and conversational APIs and protocols) also includes "**an agent architecture**" as claimed. (Schmandt, ¶127.) Coffman's "**agent architecture**" comprises the conversational APIs and protocols that enable

communication between the CVM, conversational subsystems, and conversationally-aware applications. (*Id.*) Coffman Figure 7 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, 5:1-6, 12:28-30, 13:17-28, 13:29-32, 19:16-19, 22:10-23:3, 24:10-11, Fig. 2.)

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, “**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.

Starting with the “**agent manager**,” the ’720 patent generally describes a “manager” as corresponding to software, and the “**agent manager**” as performing operations for managing agents. (’720, 3:20-28, 21:67-22:6; Schmandt, ¶129.)

Coffman similarly describes CVM software functionality to manage registration of conversationally-aware applications with the system, i.e., an “**agent manager.**” (Schmandt, ¶130.) 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; Schmandt, ¶130.)

Turning to the “**system agent**,” 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; Schmandt, ¶131.) In Coffman, the

“**system agent**” is disclosed by CVM software including the “CVM kernel.” (Schmandt, ¶131.) 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 to perform operations on behalf of both a user and conversationally-aware applications. (Schmandt, ¶131.)

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”); Schmandt, ¶132.)

Additionally, 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. (Schmandt, ¶133.) Coffman explains, for example, that I/O resources will register with the CVM kernel, and that the CVM “automatically handles all input/output issues.” (Coffman, 19:20-25, 13:5-8, 58:8-9.) Coffman further explains that the CVM’s default input/output functionality can be overridden by, for example, “redirecting” the results provided by an application. (Coffman, 41:23-26, 35:2-6; *see also id.*, 43:6-7, 44:13-16; Schmandt, ¶133.)

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, 66:8-9; Schmandt, ¶134.) 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. (Coffman, 29:1-13; Schmandt, ¶134.)

Turning to the “**agent library**,” this is disclosed by Coffman’s conversational subsystems. (Schmandt, ¶135.) 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; Schmandt, ¶135.) As 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.)

Next, Coffman teaches, and a POSITA 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**” 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. (Schmandt, ¶137.) 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**, 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, 29:1-4, 13:8-13 & 58:9-11 (CVM converts application “output into speech...provided to the user”), 44:21-22, 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**.” (Schmandt, ¶137.) 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. (*Id.*) The spoken response would qualify because the application would use text-to-speech services in order to generate the response. (*Id.*)

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**.” (Schmandt, ¶138.) Coffman explains that the CVM kernel (“**system agent**”) generally provides “output formatting and presentation.” (Coffman, 20:8-14; *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, 39:30-32.) 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.” (Schmandt, ¶138.) And 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.**” (Schmandt, ¶139.)

2. Claim 2 (2[pre])

The ’720 patent generally describes an “**event manager**” as a software module for managing events, and uses the term “**event**” consistent with the ordinary understanding as broadly referring to an action or occurrence to which a program might respond. (’720, 3:20-28, 21:61-62; Schmandt, ¶140 (citing EX1009, p.0015 (“event”)).)

A POSITA would have understood and found it obvious that Coffman’s PVA-based system (“**the mobile system**”) would include the claimed “**event manager.**” As explained below, Coffman’s “**event manager**” corresponds to software

functionality that coordinates interaction among components of its “**natural language speech processing system.**” (Schmandt, ¶141.)

(a) “sends and receives events...” (2[a])

As discussed, 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:

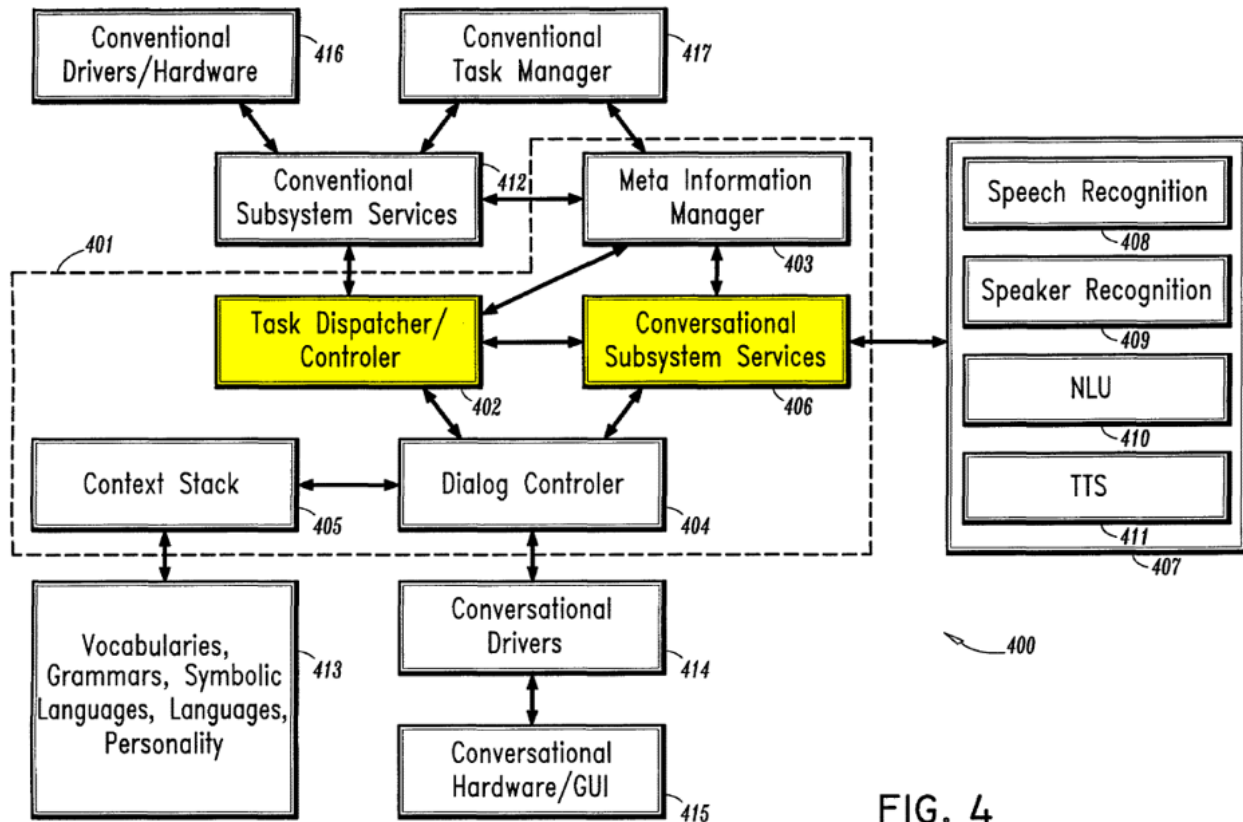


FIG. 4

(Coffman, Fig. 4 (highlighting added); Schmandt, ¶142.)

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**”). (Schmandt, ¶143.) As to task dispatcher/controller 402, Coffman explains 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...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.) As 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.)

A POSITA 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 system’s CVM, conversational subsystems, and conversational APIs (application program interfaces) and protocols (“**a plurality of components of the natural language speech processing system**”). (Schmandt, ¶144.) As discussed, 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; Schmandt, ¶144.)

(b) “provides a multi-threaded environment...” (2[b])

Task dispatcher/controller 402 and conversational subsystem services 406 (“**event manager**”) in Coffman’s system “**provide a multi-threaded**

environment.” (Schmandt, ¶145.) As discussed, Coffman teaches that task dispatcher/controller 402, among other things, manages threads/processes for the 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, 30:26-32.)

Task dispatcher/controller’s ability to provide a “**multi-threaded environment**” enables a user of Coffman’s system to simultaneously interact with multiple applications. (Schmandt, ¶146.) 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.**” (*Id.*) 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/commands. (Coffman, 42:12-13 (“Simultaneously completed tasks”), 39:5-8 (“output of multiple processes can simultaneously collide”).)

Coffman also makes clear that responses are “**real-time.**” (Schmandt, ¶146.) In the ’720 patent, “real-time” relates to a time frame allowing for natural interaction with a user. (*Id.* (citing ’720, 30:47-50).) This is consistent with the general

understanding of “real-time” in computer systems. (*Id.* (citing EX1009, p.0029 (defining “real-time”)).) Coffman’s descriptions mirror the ’720 patent and this general understanding. (Coffman, 45:29-32 (describing minimizing delay and dialog delays “minimized to acceptable levels”); Schmandt, ¶146.) Moreover, providing responses in “**real-time**” would have been obvious, since the point of Coffman’s system is to enable a user to carry on an interactive voice-based dialog. (Coffman, 11:10-12, 6:32-7:3, 56:26-57:9; Schmandt, ¶146.) The usefulness of Coffman’s system would be significantly impaired if it responded to user input only after an unreasonable delay. (Schmandt, ¶146.)

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.**” (Schmandt, ¶147.) As discussed for 1[b][ii][1], Coffman explains that its 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, 30:13-20, 5:15-20.) 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, 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”).

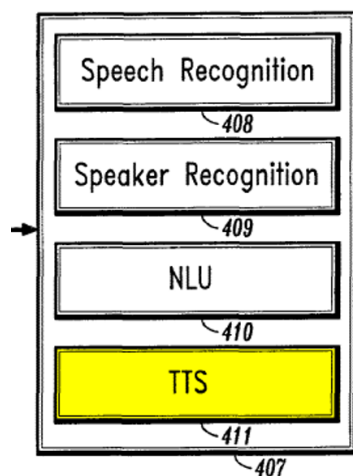
(Schmandt, ¶147.)

3. Claim 3

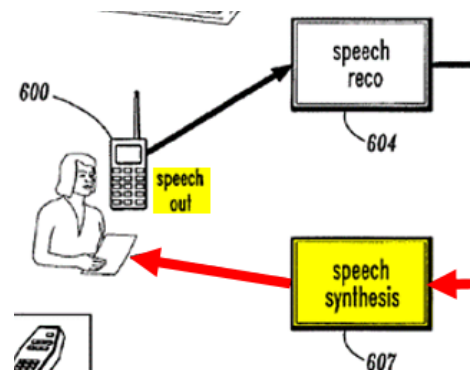
Claim 3 was addressed by 2[b] above. (Schmandt, ¶148.) Claim 3 adds that the claimed multi-threaded environment enables the natural language speech processing system “to use a plurality of the domain agents” to provide the plurality of real-time responses. As explained, task dispatcher/controller 402 enables a user of Coffman’s system to simultaneously interact with multiple applications—i.e., “to use a plurality of the domain agents.” (Coffman, 42:12-13, 39:5-8; Schmandt, ¶149.)

4. Claim 4

As explained for 1[b][iii], Coffman discloses that a conversationally-aware application may provide a text-based response (“the response include[] a text string”) that a text-to-



speech engine uses to generate a spoken response to announce to the user (“a text to speech engine that creates an encoded speech message to be announced to the user”). (Coffman, 10:28-30 (“text-to speech”), 29:1-4 (“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 (“user can listen to the output of a task”), 40:9-19, 13:8-13, 58:9-11, 39:3-5; Schmandt, ¶150.)

A POSITA would have understood that the output from the text-to-speech engine would be an “**encoded**” speech message. (Schmandt, ¶150; Coffman, 45:21-22, 26:9-27 (describing speech coders).) This is because in a speech-processing system using conventional digital computing components, such as Coffman’s system, a POSITA would have understood that, to process and output audio, the audio would be encoded in digital form. (Schmandt, ¶150 (citing Coffman, 13:1-3; EX1009, pp.005 (“analog-to-digital converter”), 0013 (“digital signal”), 0014 (“digital-to-analog converter”), 0024 (“microphone”), 0010 (“codec”)).)

5. Claim 5

It would have understood obvious that a conversationally-aware application (“**the selected domain agent**”) would include data associated with at least “**driving directions**” or “**travel information.**” (Schmandt, ¶151 (citing Coffman, 57:5-9).)

6. Claim 6

A POSITA 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.**” (Schmandt, ¶¶152-155.)

For example, Coffman describes a situation where a PVA provides driving directions to a store by fetching “the location of the store, locally or from a server.” (Coffman, 57:5-7.) A POSITA would have found it obvious that, in order for a navigation application to retrieve the store location from a server, the application would include data on how to communicate with the server, such as the server’s network address, communication protocol, and required data and input format. (Schmandt, ¶153.)

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 various ways. A POSITA would have understood that, 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. (Schmandt, ¶154.) Moreover, a POSITA would have understood that the name dialer application would include phone numbers to dial. (*Id.*) A phone number is

“**data for communicating with one or more devices**” because it enables a communication connection to be established between devices. (*Id.*) In addition, Coffman makes clear that a command to call someone using the application can involve the application directing a command 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; Schmandt, ¶154.) Like the navigation application retrieving location information from a server, a POSITA would have understood that 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. (Schmandt, ¶154.) Coffman’s teachings, and the understanding of a POSITA, with respect to the name dialer application are confirmed by Maes (EX1007), which Coffman incorporates by reference, and which provides further details about the operation of a name dialer. (Coffman, 22:25-29; Maes, 21:6-15 (describing locally-stored address book information, phone automatically dialing phone number upon recognizing command, and contacting larger remote directory for additional address book information); Schmandt, ¶155; *see also* EX1021 (Maes application corresponds to “YO999-113P”).)

7. Claim 7

As explained for claim 6, Coffman teaches that a conversationally-aware application, such as a navigation application or name dialer would include data to control another device so that it would operate in a particular way. (Schmandt, ¶156.) 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. (*Id.*)

8. Claim 8

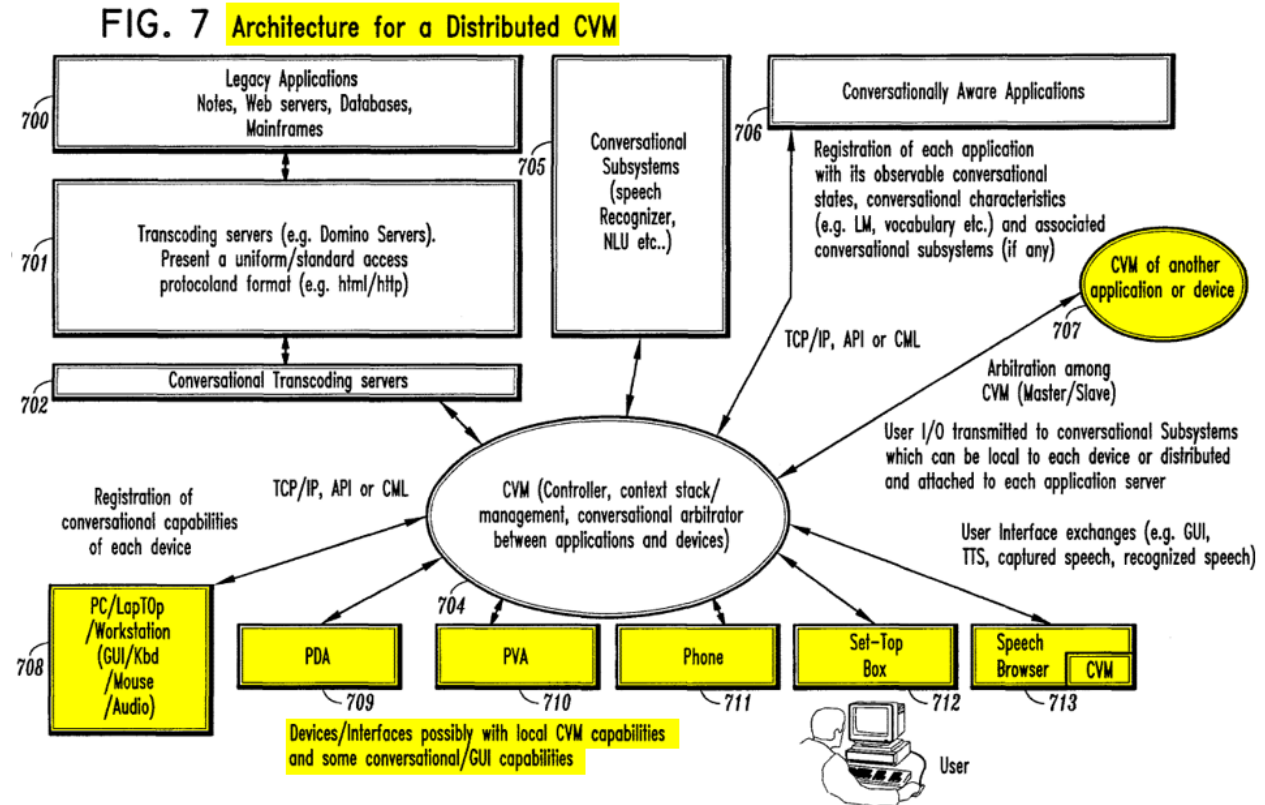
Claim 8 simply requires that one of the devices “**is connected**” to the vehicle, and does not require, for example, that the device be physically coupled to the vehicle. (Schmandt, ¶157; ’720, 3:3-5 (describing “wireless connection” to vehicle device).) In Coffman’s PVA-based system, it would have been obvious for one of the devices to be “connected” to the vehicle. (Schmandt, ¶157.) For example, as to a navigation application running on the PVA, by contacting a remote server for information, the server would be “**connected to the vehicle.**” (*Id.*) Similarly, a

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

Even if claim 8 required a physical connection to the vehicle, it would have been obvious. For example, as discussed for claim 6, a POSITA would have understood that, 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. (Schmandt, ¶158.) Coffman provides an example in which both a 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. (Coffman, 56:22-30.) In this example, it would have been obvious for the smart phone (“**at least one of the one or more devices**”) to be “**connected to the vehicle,**” since physically connecting a phone to a vehicle was routine and conventional by at least 2001. (Schmandt, ¶158 (citing EX1009, p.0018 (“hands-free kit”); EX1017, pp.003, 005, 009 (definitions of “car phone,” “cellular radio” and “transmobile”)).)

9. Claim 9

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



(Coffman, Fig. 7 (highlighting added).) As shown, the distributed architecture can involve multiple devices running CVMs. Coffman provides an example architecture where both a PVA 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; Schmandt, ¶160.)

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 PVA. (Coffman, 23:16-25 ("simultaneously control [multiple CVM or conversationally aware multiple

devices] through one single interface”).) 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.”** (Schmandt, ¶161.)

Similarly, Coffman contemplates that, with a desktop PC acting as the master CVM for example, voice input at the PVA could be transmitted to the desktop PC for processing by the desktop PC’s CVM and associated conversational subsystems (**“natural language speech processing system”**). (Schmandt, ¶162; Coffman, 25:27-26:1, 26:4-6; *see also id.*, 26:6-25 (describing speech transmission protocols).) Therefore, voice input to a speech unit at a PVA could be transmitted to a desktop PC conversational system for processing, and the results returned to the PVA. 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.”** (Schmandt, ¶162.)

10. Claim 10

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 (“the location of the store is fetched, locally or from a server”).) Therefore, Coffman discloses **“wherein the at least one device connected to the vehicle includes...a navigation system.”** (Schmandt, ¶163.)

11. Claim 11

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; Schmandt, ¶164 (citing EX1009, p.0034 (“telematics”)).) A “**telematics control unit**” simply refers to a device that controls telematics functionality. (Schmandt, ¶164 (citing EX1009, p.0012 (“control unit”)).) By 2001-2002, “**telematics control unit[s]**” were well-known and in commercial use, and thus neither inventive nor non-obvious. (Schmandt, ¶164 (citing EX1009, p.0034 (“telematics”); EX1017, pp.006-007 (“telematics”)).)

In Coffman, the claimed “**telematics control unit**” corresponds to the vehicle’s PVA. (Schmandt, ¶165.) 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, 56:24-25.) 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.)

As explained for claim 8, Coffman discloses that the vehicle’s PVA is also connected to the user’s smart phone in the vehicle and to 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; Schmandt, ¶166.)

12. Claim 12

As explained for claim 1, a POSITA would have understood and found it obvious that Coffman’s PVA would include a **“speech unit”** and/or a **“natural language speech processing system.”** (Schmandt, ¶167 (citing Coffman, 56:21-25, 55:27-28).) 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”**). (Schmandt, ¶167 (citing Coffman, 13:1-3, 29:1-13, 55:27-28, 56:21-25).)

13. Claim 13

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. (Coffman, 57:5-7 (remote server provides

location/navigation service), 42:2-6 (remote machine providing contact information service).) 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.)

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.**” (Schmandt, ¶¶168-169.)

14. Claim 14

As discussed for claim 6, Coffman teaches a remotely located service that includes “**a location service.**” (Schmandt, ¶170 (citing Coffman, 57:5-7).)

15. Claim 15

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 PVA 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, 26:4-6, 26:6-25 (describing speech transmission protocols); Schmandt, ¶171.) A POSITA 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. (Schmandt, ¶171 (citing Coffman, 63:3-6 (describing “connecting via a PSTN 1002 and internet/intranet 1003 networks”).) It would have been obvious, therefore, that a PVA, acting as a client of a master CVM at the user’s desktop PC, would communicate with it via a communication network. (Schmandt, ¶171.)

A POSITA would have also understood and found it obvious that the “**speech unit**” of the PVA and the “**natural language processing system**” of the desktop PC would “**each include respective transceivers**” for communicating with each other. (Schmandt, ¶172.) A “**transceiver**” is a conventional and generic device for transmitting and receiving data. (*Id.* (citing EX1009, p.0036 (“transceiver”); EX1017, p.008 (“transceiver”))).) 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.**” (*Id.*)

16. Claim 16

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.**” (Schmandt, ¶173.)

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; Schmandt, ¶173 (citing EX1009, p.0019 (“internetwork” entry)).) And because the communication involves a device in a mobile vehicle, a POSITA 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. (Schmandt, ¶173 (citing EX1017, p.004 (“cellular modem”)).) 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. (*Id.* (citing Coffman, 63:14).)

17. Claim 22

Claim 22 is substantially the same as claim 9 for purposes of this Petition and would have been obvious for the same reasons. (Schmandt, ¶174.)

18. Claim 23

It would have been obvious for Coffman’s PVA-based conversational system to comprise “**at least one of a display or a keypad**.” (Schmandt, ¶175.) Coffman makes clear that the conversational techniques incorporate and improve devices with displays and/or keypads. (Coffman, 11:29-30, 12:8-10; Schmandt, ¶175.) Including a display and/or keypad with a PVA would beneficially provide the user with additional ways of inputting and receiving information. (Schmandt, ¶175.) 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 POSITA 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. (Schmandt, ¶175 (citing EX1009, p.0034 (“telematics” entry describing “dashboard navigation systems”)).)

19. Claim 24

Claim 24 is substantially the same as claim 12 for purposes of this Petition. (Schmandt, ¶176.) 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 PVA, which Coffman teaches could in turn be “**embedded in...the vehicle**” or “**a fixed device connected to the vehicle.**” (*Id.* (citing Coffman, 12:23-25 (“car embedded computer”), 23:23-25 (same), 55:27-28.)

20. Claim 25

It would have been obvious for “the communicatively coupled services” to “**include at least one shared network resource.**” As discussed, 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. (Schmandt, ¶177.) This is confirmed by the Maes, which further describes a name dialer application, explaining that the contact information service can be provided by “a larger corporate (or public) directory.” (Schmandt, ¶177 (citing Maes, 21:12-15).)

21. Claim 27

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**.” (Schmandt, ¶¶178-181.)

As to “**text and string handling**,” a POSITA would have understood that Coffman’s conversational subsystems perform text and string handling in numerous ways. (Schmandt, ¶179.) 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.”⁴ (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

⁴ “ASCII input” refers to text formatted using the American Standard Code for Information Interchange (“ASCII”). (EX1009, p.006 (“ASCII”).)

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 POSITA would have understood that these transcription, parsing and translating operations would involve “text and string handling” functions. (Schmandt, ¶179.)

As to “**fuzzy and probabilistic evaluation**,” a POSITA 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. (Schmandt, ¶180.) 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 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. (Schmandt, ¶180.)

As 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. (Schmandt, ¶181; Coffman,

47:8-13 (“the personality/behavior of the CVM (e.g. synthesized voice - Voice Fonts)”), 8:1-3 (“Voice fonts, which allows the user to select the type of voice providing the output”), 39:30-32.)

22. Claim 29 (29[pre])

Coffman’s “**mobile system**” includes speech processing functionality, including speech and speaker recognition, disclosing the “**speech recognition engine.**” (Schmandt, ¶182 (citing Coffman, Fig. 4, 29:1-4).) It would have been obvious, as explained below, for Coffman’s “**speech recognition engine**” to perform the claimed steps.

(a) “determines...” (29[a])

Coffman discloses “text-independent speaker identification” (Coffman, 47:25-27, 8:16-18, 46:9-12), a form of speaker identification based on the user’s unique voice characteristics. (Schmandt, ¶183 (citing EX1008, pp.0047 (“speaker recognition”), 0051 (“text-independent speaker recognition system”).) Kanevsky, incorporated by Coffman (Coffman, 37:5-13), further confirms 29[a] is disclosed and obvious. (Schmandt, ¶183 (citing Kanevsky, 6:66-7:8, 6:12-15, 2:6-8).)

(b) “tags...” (29[b])

“Tagging” is generally understood as creating an association between a piece of information and another piece of information, such as a characteristic or property. (Schmandt, ¶184 (citing ’720, 27:5-12).) A POSITA would have understood and

found it obvious that Coffman’s “**speech recognition engine**,” which both recognizes words and phrases in an utterance and determines user identity, would “**tag**” the recognized words and phrases with the user identity. (Schmandt, ¶185.) Coffman makes clear that an association is created between the recognized words and phrases and user identity, explaining that “user identity can be associated in the context of each discourse.” (*Id.* (citing Coffman, 48:1-3, 46:26-31, 48:11-14).) A POSITA 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/discourse for a current dialog that best match the user’s recognized words and phrases; and accordingly 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.**” (*Id.*)

23. Claim 30

The analysis for 1[b][i] addressed claim 30. 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**” 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. (Schmandt, ¶186.)

24. Independent Claim 31: “A method...” (31[pre])

The preamble of claim 31 is substantially the same as the preamble of claim 1 for purposes of this Petition; as explained below, the claimed method is disclosed by and obvious over the prior art. (Schmandt, ¶¶187-203.)

(a) “receiving...” (31[a])

31[a] recites a method counterpart to 1[a], and is therefore disclosed and obvious as discussed for 1[a]. (Schmandt, ¶188.)

(b) “recognizing...” (31[b])

31[b] is largely a method counterpart 1[b][i]. (Schmandt, ¶189.) 31[b] recites that the speech recognition engine is “**connected to the computer device on the vehicle.**” The analysis of 1[b] explained that the claimed “natural language speech processing system” (of which the speech recognition engine is a part) would be connected to the PVA (“**connected to the computer device on the vehicle**”). (Schmandt, ¶190.) Therefore, for the reasons discussed for 1[b] and 1[b][i], 31[b] is disclosed and obvious. (Schmandt, ¶191.)

(c) “determining...” (31[c])

31[c] is largely a method counterpart to 1[b][ii] and 1[b][ii][1]. (Schmandt, ¶192.) 31[c] further recites that the parser is “**connected to the computer device**

on the vehicle.” The analysis of 1[b] explained that the claimed “natural language speech processing system” (of which the parser is a part) would be connected to the PVA (“**connected to the computer device on the vehicle**”). (Schmandt, ¶193.) Therefore, for the reasons discussed for 1[b], 1[b][ii] and 1[b][ii][1], 31[c] is disclosed and obvious. (Schmandt, ¶194.)

(d) “**selecting...**” (31[d])

(e) “**transforming...**” (31[e])

31[d] is largely a method counterpart to 1[b][ii] and 1[b][ii][2], and 31[e] is largely a method counterpart to 1[b][ii] and 1[b][ii][3]. (Schmandt, ¶¶195, 198.) Both also recite that the parser is “**connected to the computer device on the vehicle,**” which was addressed in the analysis of 1[b]. (Schmandt, ¶¶196, 199.) Therefore, for the reasons discussed for 1[b], 1[b][ii], 1[b][ii][2] and 1[b][ii][3], 31[d] and 31[e] are disclosed and obvious. (Schmandt, ¶¶197, 200.)

(f) “**forwarding...**” (31[f])

31[f] is largely a method counterpart to 1[b][iii]. (Schmandt, ¶201.) 31[f] further recites that the agent architecture is “**connected to the computer device on the vehicle.**” This was addressed in the analysis of 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 PVA (“**connected to the computer device on the vehicle**”). (Schmandt, ¶202.)

31[f] further recites “**forwarding the formulated question or command**” to the agent architecture, which would have been obvious. (Schmandt, ¶203.) 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, 10:28-30, 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 “**forward**[ed]” 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. (Schmandt, ¶203.)

25. Claim 35

Claim 35 is essentially an amalgam of claims 6, 8 and 10. (Schmandt, ¶204.) As discussed for those claims, it would have been obvious for Coffman’s 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. (Schmandt, ¶204.)

26. Claims 36-37

Claims 36 and 37 are not meaningfully different from claims 5 and 13, respectively, and therefore would have been obvious for the reasons discussed for those claims. (Schmandt, ¶¶205-206.)

27. Claim 38

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. (Schmandt, ¶207 (citing Coffman, 57:5-7).) More generally, Coffman describes how a CVM device can access information across a network. (*Id.* (citing Coffman, 63:3-6).) 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. (*Id.*)

28. Claim 39

A POSITA 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. (Schmandt, ¶208 (citing Coffman, 57:5-7, 63:3-6).) The service and server would be

“**associated**” because the server provides the service and moreover, a POSITA would have understood that information associating the service with the server (such as an IP address or URL) would enable a CVM device to communicate with the service provided by the server. (*Id.*)

29. Claim 40

As discussed for claim 38, Coffman discloses that a device can communicate with a remotely located service via a communication network. (Schmandt, ¶209 (citing Coffman, 63:3-6).) A PVA’s communication with, for example, the remotely located location/navigation service would take place “**via a communication network**” would have therefore been obvious. (*Id.*)

30. Claim 41

Claim 41 is not meaningfully different from claim 14 and therefore would have been obvious for the same reasons. (Schmandt, ¶210.)

31. Claim 42

As discussed for claim 38, Coffman describes transmitting a request from a vehicle to a remotely located location/navigation service. (Schmandt, ¶211 (citing 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 in a smart phone and/or cell tower, because it would have involved communicating with a remote service from a highly mobile environment, a situation routinely addressed

using cellular devices and networks. (*Id.* (citing Coffman, 56:26; EX1009, p.0032 (“smartphone”); EX1017, p.004 (“cellular radio”))).)

32. Claim 43

Claim 43 is not meaningfully different from claim 25 and therefore would have been obvious for the same reasons. (Schmandt, ¶212.)

33. Claims 52, 55

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.) 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. (Schmandt, ¶213.) A POSITA 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. (*Id.*)

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) (i.e., claim 52). (Schmandt, ¶213.) Coffman also therefore discloses providing “**interactive guidance**” (turn-by-turn guidance) “**relating to one or more tasks for the user**” (driving to a store) “**based**

on a location of the vehicle” (using the vehicle’s location) (i.e., claim 55).
(Schmandt, ¶214.)

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

1. Claim 18

Turning first to “**speech coder**,” a “**coder**” is simply functionality that can convert an analog audio or video signal to a digital form. (Schmandt, ¶215 (citing EX1009, p.0010 (“codec”))). A POSITA 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. (*Id.*)

In fact, Coffman expressly discloses that its devices include a “**speech coder**.” Referring to Figure 3 (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:

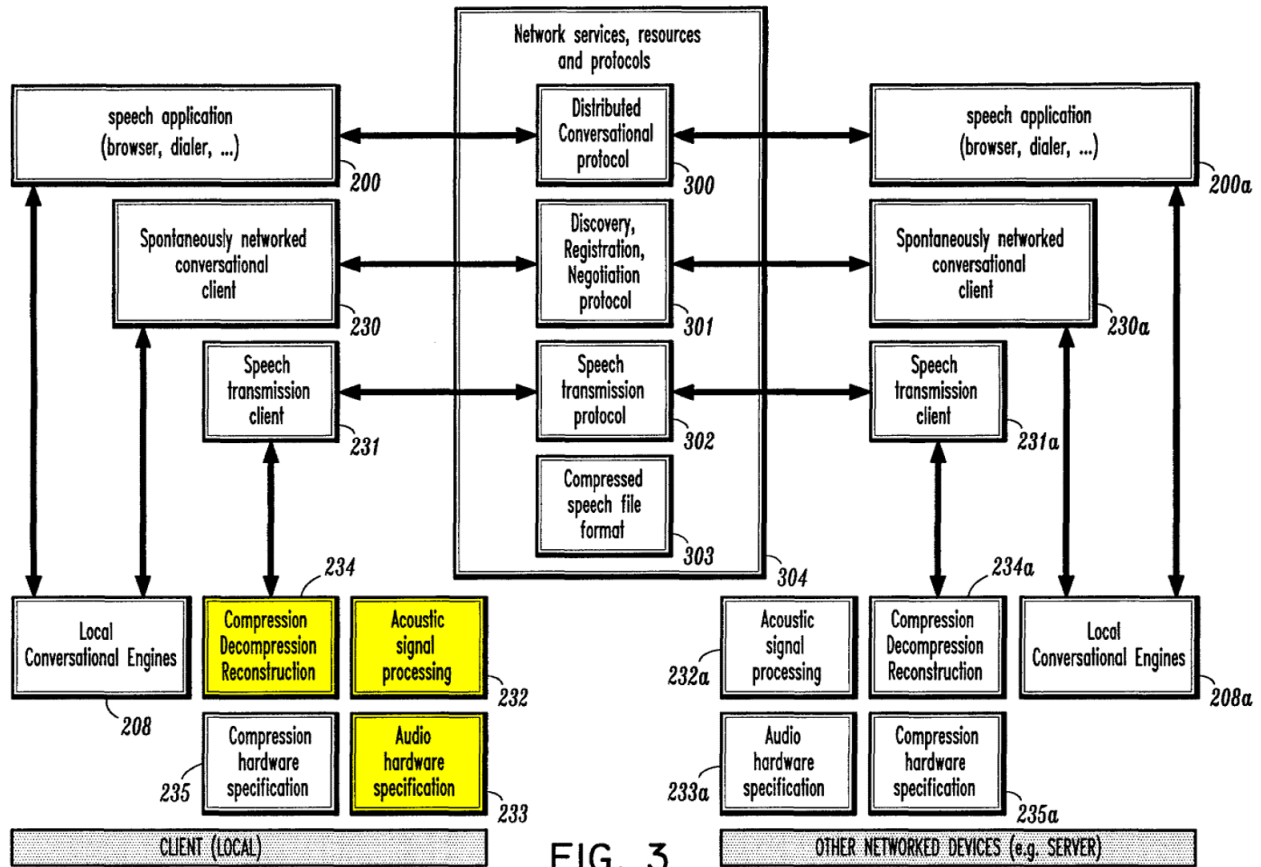


FIG. 3

(Coffman, Fig. 3 (highlighting added), 23:3-5, 26:9-16, 26:16-27; *see also id.*, Fig. 7 (illustrating distributed architecture).) A POSITA 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 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.”** (Schmandt, ¶216.)

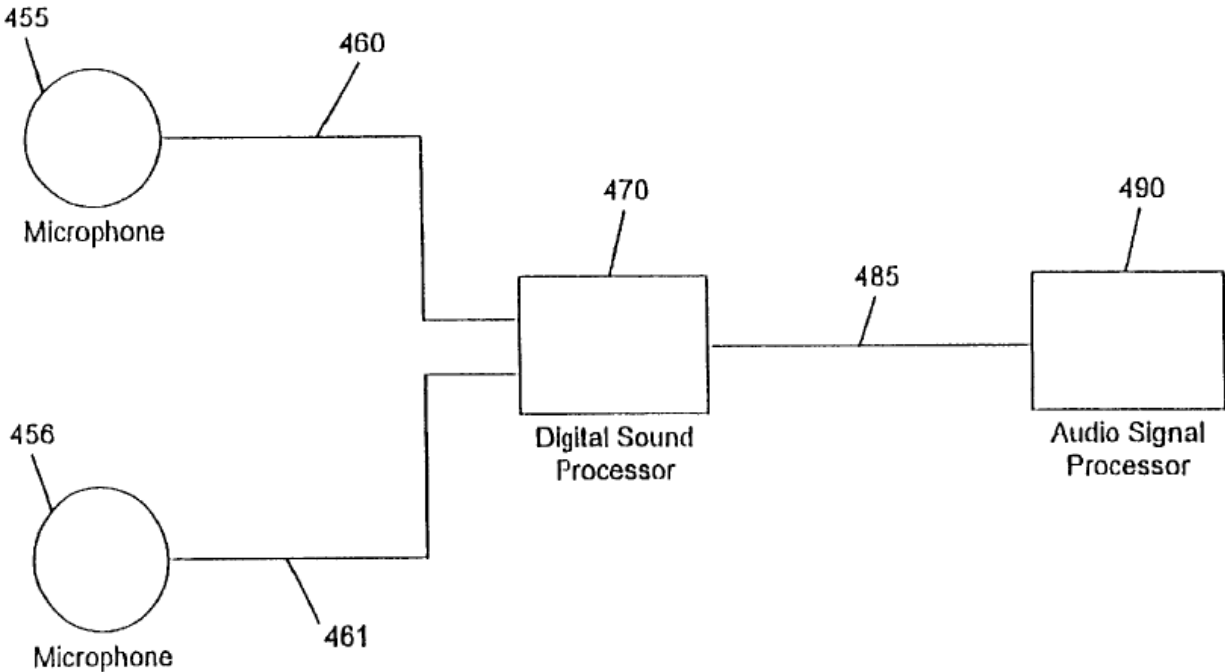
As to **“an array microphone that receives the natural language speech**

utterance,” Coffman does not go into detail about microphone arrangements in its conversational system. (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, 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 **DeLine (EX1010)**. (Schmandt, ¶217.)

DeLine 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; Schmandt, ¶218.) 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 and that “[t]his may be especially troublesome when the other party is an automated voice system which responds to a particular voice communication.” (DeLine, 1:40-2:33, 2:29-33, 2:18-23; ; *see generally id.*, 2:15-33.)

In an embodiment, DeLine proposes a microphone array as illustrated in Figure 12 below:

450



(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 touts the advantages of using input from multiple microphones, explaining that 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;...adaptive signal processing techniques and adaptive noise cancellation techniques.” (DeLine, 48:26-45.)

Turning back to the claim language, an “**array microphone**” is a device comprising multiple microphones. (Schmandt, ¶220.) DeLine’s use of two or more microphones in this manner discloses “**an array microphone that receives the natural language speech utterance.**” (*Id.* (citing DeLine 48:53-63 (describing “array of multiple mirror-mounted microphones” and commercially-available microphone array from Andrea Electronics Corporation))).) As noted, during ’720 patent prosecution, the Examiner cited DeLine for its disclosures related to this limitation, which the applicant did not dispute. (EX1019, pp.00357-00359, 00412-00413, 00441-00442.)

As to the claimed “**filter**,” this is disclosed by DeLine’s discussion of system components to perform various noise-cancelling techniques, as discussed. (Schmandt, ¶221 (citing 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. (*Id.* (citing DeLine, e.g., 1:20-27 (speech signal “with an enhanced vocal content and a reduced ambient noise content”))).)

Rationale and Motivation to Combine (Ground 1 Prior Art and DeLine).

It would have been obvious to combine DeLine with Coffman and its conversational system under the proposed combination, predictably resulting in the speech unit of a PVA in Coffman’s system further including multiple microphones and components for performing noise-cancelling operations on the electronic signals provided from the microphones. (Schmandt, ¶222.) Combining DeLine and Coffman in this manner would have resulted in a speech unit that included “**an array microphone that receives the natural language speech utterance**” and “**a filter that optimizes a signal to noise ratio of the electronic signal.**” (*Id.*)

DeLine is an analogous reference and would have been reasonably pertinent to problems facing the patent inventors. (Schmandt, ¶223; DeLine, 1:20-34, 2:15-33.)

A POSITA would have been motivated to combine. (Schmandt, ¶224.) Coffman describes using its 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 POSITA would have been motivated to incorporate them into the similar environment contemplated by Coffman for the same benefits. (Schmandt, ¶224.)

A POSITA would have had a reasonable expectation of success. (Schmandt, ¶225 (citing DeLine, 47:32-41, 48:30-45, 48:53-63).)

2. Claim 20

A “**one-dimensional**” microphone array is simply a linear array. (’720, 25:48-53.) An arrangement of two (or more) microphones, as DeLine describes, would be “**at least a one-dimensional**” (or linear) array. (Schmandt, ¶226.)

3. Claim 21

As discussed for claim 18, Coffman discloses a “**speech coder**” that uses “**audio compression**” for encoding a user’s utterance as an electronic signal. (Schmandt, ¶227 (citing Coffman, Fig. 3, 26:4-27).)

It would have been obvious to use “**adaptive lossy audio compression**,” a well-known compression type that a POSITA would have recognized was particularly suited to processing and transmitting voice data in a mobile system. (Schmandt, ¶228.) For example, a POSITA 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. (*Id.* (citing EX1009, pp.0023 (“lossy compression”), 0011 (“compress”)).) Moreover,

the **Maes** application (**EX1007**) that Coffman incorporates and cites for additional information about encoding speech (Coffman, 22:25-29, 26:23-24) makes clear that a POSITA would have recognized that “any suitable compression scheme” for encoding speech could be used. (Schmandt, ¶228 (citing Maes, 16:5-6).) Maes encourages use of a compression rate that “can be adapted between robustness (error rate - feature distortion) and perceptual quality.” (*Id.* (citing Maes, 16:9-15).) Maes therefore confirms that a POSITA would have considered “**adaptive lossy audio compression**” to be a “suitable compression scheme” and obvious choice for Coffman’s system. (*Id.*)

4. Claim 32

Claim 32 is substantially identical to claim 18, and would have been obvious for the same reasons. (Schmandt, ¶229.)

5. Claim 33

As discussed for claim 18, DeLine discloses 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, 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 to include a filter that “**removes background**

noise from the electronic signal to optimize the signal to noise ratio of the electronic signal.” (Schmandt, ¶230.)

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

1. Claim 19

As discussed for claim 18, it would have been obvious in view of DeLine for the speech unit of a PVA to include a filter employing various noise-cancelling techniques. (DeLine, 48:30-45.) For 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, DeLine identifies “adaptive signal processing techniques and adaptive noise cancellation techniques” (DeLine, 48:38-40), which suggests that echo cancellation would include adaptive techniques. (Schmandt, ¶231.) In any event, “adaptive echo cancellation” would have been obvious over Gerson (EX1011). (*Id.*)

Gerson relates to a “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; Schmandt, ¶232.) 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, 1:35-40.) 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.)

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 (and incorporates) U.S. Patent No. 5,561,668 as an example. (Gerson, 9:26-34; *see generally id.*, 9:24-10:10 (describing echo cancellation).)

A POSITA would have understood that Gerson discloses “**adaptive echo cancellation**.” (Schmandt, ¶234.) To start, the Examiner cited Gerson during ’720 patent prosecution for its disclosures related to this limitation, which the applicant did not dispute. (EX1019, pp.00357-00359, 00413, 00441-00442.) A POSITA would have understood Gerson’s echo cancellation to be “**adaptive**” because, as the Examiner noted, it eliminated echoes and noise in varying conditions. (Schmandt, ¶234.) Moreover, a POSITA would have understood that Gerson’s echo cancellation technique—“acoustic echo cancellation”—is “**adaptive**.” (Schmandt, ¶235 (citing EX1012, p.0016 §3.2.2 (“[a]coustic echo cancellation works by forming in an

adaptive manner a replica of the echo path....[T]he filter has to be adaptive”);
EX1013, 4:7-10, Figs. 3 & 4).)

Rationale and Motivation to Combine (Ground 2 Prior Art and Gerson).

It would have been obvious to combine Gerson with Coffman’s conversational system under the proposed combination, predictably resulting in the PVA’s speech unit in Coffman’s system as taught by the combination of Coffman and DeLine further using “**adaptive echo cancellation.**” (Schmandt, ¶236.) Combining in this manner would have 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. (*Id.*)

Gerson is an analogous reference and would have also been reasonably pertinent to problems facing the patent inventors. (Schmandt, ¶237; Gerson, 1:8-13.)

A POSITA would have been motivated to combine. (Schmandt, ¶238.) As noted, Gerson explains it is “human nature” to barge-in during presentation of audio output. (Gerson, 1:35-39.) Gerson’s use of acoustic echo cancellation, by allowing barge-in, advantageously enables a more natural and efficient interaction with a speech-based system. (Schmandt, ¶238.) Because Coffman likewise aims to provide a conversational system where a user and the system interact using speech/audio, a POSITA would have appreciated the applicability of Gerson’s

teachings and benefits of incorporating them. (*Id.*) Moreover, as to the speech unit in Coffman’s PVA, Gerson’s techniques can be applied “with particular advantage” to in-vehicle systems (Gerson, 6:10-12), further confirming and motivating the combination. (Schmandt, ¶238.) A POSITA therefore would have been motivated to incorporate acoustic echo cancellation into the similar environment contemplated by Coffman to obtain the same benefits. (*Id.*)

A POSITA would have had a reasonable expectation of success, as it would have involved applying conventional techniques and technologies without undue experimentation. (Schmandt, ¶239; Gerson, 9:26-34.)

2. Claim 34

Claim 34 is not meaningfully different from claim 19 and would have been obvious for the same reasons. (Schmandt, ¶240.)

VII. CONCLUSION

Petitioner respectfully requests IPR institution.

Dated: April 5, 2024

Respectfully submitted,

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004
Tel: (650) 843-5001
Fax: (650) 849-7400

By: / Heidi L. Keefe /
Heidi L. Keefe
Reg. No. 40,673
Counsel for Petitioner

CERTIFICATE OF COMPLIANCE WITH WORD COUNT

Pursuant to 37 C.F.R. § 42.24(d), I certify that this petition complies with the type-volume limits of 37 C.F.R. § 42.24(a)(1)(i) because it contains 13,987 words, according to the word-processing system used to prepare this petition, excluding the parts of this petition that are exempted by 37 C.F.R. § 42.24(a) (including the table of contents, a table of authorities, mandatory notices, a certificate of service or this certificate word count, appendix of exhibits, and claim listings).

DATED: April 5, 2024

COOLEY LLP
ATTN: Patent Docketing
1299 Pennsylvania Avenue NW
Suite 700
Washington, D.C. 20004
Tel: (650) 843-5001
Fax: (650) 849-7400

/ Heidi L. Keefe /
Heidi L. Keefe
Reg. No. 40,673

CERTIFICATE OF SERVICE

I hereby certify, pursuant to 37 C.F.R. Sections 42.6 and 42.105, that a complete copy of the attached **PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 7,693,720 B2**, including all exhibits (**Nos. 1001-1021**) and related documents, are being served via Federal Express on the 5th day of April, 2024, the same day as the filing of the above-identified document in the United States Patent and Trademark Office/Patent Trial and Appeal Board, upon Patent Owner by serving the correspondence address of record with the USPTO as follows:

174899 - The Law Offices of David A. Gerasimow, P.C.
211 W. Wacker Dr.
Ste. 1717
Chicago, IL 60606
UNITED STATES

And, via Federal Express upon counsel of record for Patent Owner in the litigation pending before the U.S. District Court for the District of Delaware entitled *Dialect, LLC v. Google LLC*, Case No. 1:23-cv-00378 (D. Del.) as follows:

Garland Stephens
BLUE PEAK LAW GROUP LLP
3139 W Holcombe Blvd, PMB 8160
Houston, TX 77025
Telephone: 281-972-3036
garland@bluepeak.law

DATED: April 5, 2024

/ Heidi L. Keefe /
Heidi L. Keefe
Reg. No. 40,673
COOLEY LLP
1299 Pennsylvania Ave. NW
Suite 700
Washington, D.C. 20004
Tel: (650) 843-5001