

**UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

MICROSOFT CORPORATION,

Petitioner,

v.

DIALECT, LLC,

Patent Owner.

---

IPR2025-00659

U.S. Patent No. 9,495,957

---

**PETITION FOR *INTER PARTES* REVIEW OF**

**U.S. PATENT NO. 9,495,957**

## TABLE OF CONTENTS

|                                                                                                                       | <b>Page</b> |
|-----------------------------------------------------------------------------------------------------------------------|-------------|
| EXHIBIT LIST .....                                                                                                    | VI          |
| TABLE OF ABBREVIATIONS AND CONVENTIONS .....                                                                          | VIII        |
| I. INTRODUCTION .....                                                                                                 | 1           |
| II. MANDATORY NOTICES .....                                                                                           | 1           |
| A. Real Party-in-Interest .....                                                                                       | 1           |
| B. Related Matters.....                                                                                               | 1           |
| C. Lead and Back-Up Counsel and Service Information .....                                                             | 2           |
| III. PAYMENT OF FEES .....                                                                                            | 2           |
| IV. REQUIREMENTS FOR <i>INTER PARTES</i> REVIEW .....                                                                 | 2           |
| A. Grounds for Standing .....                                                                                         | 2           |
| B. Identification of Challenged Claims .....                                                                          | 3           |
| V. THE 957 PATENT.....                                                                                                | 3           |
| A. Effective Filing Date .....                                                                                        | 3           |
| B. Person of Ordinary Skill in the Art.....                                                                           | 3           |
| C. Overview of the 957 Patent.....                                                                                    | 4           |
| VI. CLAIM CONSTRUCTION .....                                                                                          | 6           |
| VII. PRINCIPAL PRIOR ART .....                                                                                        | 7           |
| A. Summary of Maes .....                                                                                              | 7           |
| B. The Combination of Maes’s Teachings with those of Coffman.....                                                     | 8           |
| VIII. THE CHALLENGED CLAIMS ARE UNPATENTABLE .....                                                                    | 13          |
| A. Ground 1: Maes renders obvious claims 7 and 8 .....                                                                | 13          |
| 1. Claim 7.....                                                                                                       | 13          |
| a. [7.0] A computer-implemented method of<br>processing a natural language utterance, the method<br>comprising: ..... | 13          |
| b. [7.1] generating a context stack comprising context<br>information that corresponds to a plurality of prior        |             |

|    |                                                                                                                                                                                                                                                                                                         |    |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
|    | utterances, wherein the context stack includes a plurality of context entries;.....                                                                                                                                                                                                                     | 18 |
| c. | Therefore, <u>Maes</u> teaches or suggests this claim element for this additional reason. Ex.1003, ¶84. [7.2] receiving, by one or more computer processors, the natural language utterance, wherein the natural language utterance is associated with a command or is associated with a request; ..... | 24 |
| d. | [7.3] determining one or more words of the natural language utterance by performing speech recognition on the natural language utterance;.....                                                                                                                                                          | 28 |
| e. | [7.4.0] identifying, from among the plurality of context entries, one or more context entries that correspond to the one or more words, wherein the context information includes the one or more context entries,.....                                                                                  | 36 |
| f. | [7.4.1] wherein identifying the one or more context entries comprises: comparing the plurality of context entries to the one or more words; .....                                                                                                                                                       | 43 |
| g. | [7.4.2] [wherein identifying the one or more context entries comprises:] generating, based on the comparison, one or more rank scores for individual context entries of the plurality of context entries; and.....                                                                                      | 45 |
| h. | [7.4.3] [wherein identifying the one or more context entries comprises:] identifying, based on the one or more rank scores, the one or more context entries from among the plurality of context entries; and .....                                                                                      | 48 |
| i. | [7.5] determining based on the determined one or more words and the context information, the command or the request associated with the natural language utterance. ....                                                                                                                                | 49 |
| 2. | Claim 8.....                                                                                                                                                                                                                                                                                            | 54 |
| a. | [8.0] The method of claim 7, further comprising: providing a prompt regarding the command or request, wherein the prompt is a prompt for additional information or a prompt for confirmation; and.....                                                                                                  | 54 |

|     |                                                                                                                                                                                                        |    |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| b.  | [8.1] receiving, in response to providing the prompt, a non-speech input, wherein the command or request is determined further based on the non-speech input. ....                                     | 56 |
| B.  | Ground 2: Maes and Coffman render obvious claims 7 and 8.....                                                                                                                                          | 57 |
| C.  | Ground 3: Kennewick and Weissman, in further view of Maes renders obvious claims 7 and 8. ....                                                                                                         | 58 |
| 1.  | Claim 7.....                                                                                                                                                                                           | 58 |
| a.  | Dialect is Collaterally Estopped from Relitigating the Obviousness of Claim 7.....                                                                                                                     | 58 |
| 2.  | Claim 8.....                                                                                                                                                                                           | 59 |
| a.  | [8.0] The method of claim 7, further comprising: providing a prompt regarding the command or request, wherein the prompt is a prompt for additional information or a prompt for confirmation; and..... | 59 |
| b.  | [8.1] receiving, in response to providing the prompt, a non-speech input, wherein the command or request is determined further based on the non-speech input. ....                                     | 59 |
| IX. | THE BOARD SHOULD NOT EXERCISE ITS DISCRETION TO DENY INSTITUTION .....                                                                                                                                 | 62 |
| A.  | The Board Should Not Deny Institution Under 35 U.S.C. § 325(d) .....                                                                                                                                   | 62 |
| B.  | The Board Should Not Deny Institution Under 35 U.S.C. § 314(a).....                                                                                                                                    | 64 |
| 1.  | Factor 1: Stay .....                                                                                                                                                                                   | 64 |
| 2.  | Factor 2: Trial Date .....                                                                                                                                                                             | 65 |
| 3.  | Factor 3: Parallel Proceeding.....                                                                                                                                                                     | 65 |
| 4.  | Factor 4: Overlapping Issues .....                                                                                                                                                                     | 66 |
| 5.  | Factor 5: Overlapping Parties .....                                                                                                                                                                    | 66 |
| 6.  | Factor 6: Strength/Other Considerations .....                                                                                                                                                          | 66 |
| X.  | CONCLUSION.....                                                                                                                                                                                        | 67 |

|                                 |    |
|---------------------------------|----|
| CERTIFICATE OF COMPLIANCE ..... | 69 |
| CERTIFICATE OF SERVICE .....    | 70 |
| CLAIM LISTING .....             | 72 |

**TABLE OF AUTHORITIES**

|                                                | <b>Page(s)</b> |
|------------------------------------------------|----------------|
| <b>Statutes</b>                                |                |
| 35 U.S.C. §103 .....                           | 1, 18          |
| 35 U.S.C. § 314(a) .....                       | 7              |
| 35 U.S.C. § 325(d) .....                       | 7              |
| United States Code Title 18 section 1001 ..... | 77             |
| <b>Other Authorities</b>                       |                |
| 37 C.F.R. § 1.68 .....                         | 1              |
| 37 C.F.R. § 42.6(e) .....                      | 10             |
| 37 C.F.R. § 42.24 .....                        | 9              |

**EXHIBIT LIST**

| <b>No.</b> | <b>Exhibit Description</b>                                                                                                                                           |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001       | U.S. Patent No. 9,495,957                                                                                                                                            |
| 1002       | File History of U.S. Patent No. 9,495,957                                                                                                                            |
| 1003       | Declaration of Dr. Henry Houh                                                                                                                                        |
| 1004       | CV of Dr. Henry Houh                                                                                                                                                 |
| 1005       | U.S. Patent Pub. No. 6,964,023 (“ <u>Maes</u> ”)                                                                                                                     |
| 1006       | RESERVED                                                                                                                                                             |
| 1007       | Second Amended Complaint                                                                                                                                             |
| 1008       | RESERVED                                                                                                                                                             |
| 1009       | RESERVED                                                                                                                                                             |
| 1010       | RESERVED                                                                                                                                                             |
| 1011       | RESERVED                                                                                                                                                             |
| 1012       | International Patent Application Publication No. WO 00/20962 (“ <u>Coffman</u> ”) filed as PCT/US99/22927                                                            |
| 1013       | U.S. Patent No. 7,137,126 (“ <u>Coffman</u> 126”)                                                                                                                    |
| 1014       | RESERVED                                                                                                                                                             |
| 1015       | RESERVED                                                                                                                                                             |
| 1016       | RESERVED                                                                                                                                                             |
| 1017       | Excerpts from Dafydd Gibbon et al., “Handbook of Multimodal and Spoken Dialogue Systems : Resources, Terminology, and Product Evaluation” (2000) (“ <u>Gibbon</u> ”) |
| 1018       | RESERVED                                                                                                                                                             |

|      |                                                                 |
|------|-----------------------------------------------------------------|
| 1019 | RESERVED                                                        |
| 1020 | RESERVED                                                        |
| 1021 | RESERVED                                                        |
| 1022 | RESERVED                                                        |
| 1023 | RESERVED                                                        |
| 1024 | RESERVED                                                        |
| 1025 | U.S. Patent No. 6,453,315 (“ <u>Weissman</u> ”)                 |
| 1026 | RESERVED                                                        |
| 1027 | RESERVED                                                        |
| 1028 | RESERVED                                                        |
| 1029 | EDTX Calendar, Judge Gilstrap                                   |
| 1030 | RESERVED                                                        |
| 1031 | RESERVED                                                        |
| 1032 | U.S. Patent Publication No. 2004/0044516 (“ <u>Kennewick</u> ”) |

**TABLE OF ABBREVIATIONS AND CONVENTIONS**

| <b>Abbreviation</b>      | <b>Meaning</b>                      |
|--------------------------|-------------------------------------|
| 957 Patent               | Ex.1001: U.S. Patent No. 9,495,957  |
| IPR                      | <i>inter partes</i> review          |
| Petitioner               | Microsoft Corporation (“Microsoft”) |
| Patent Owner or PO       | Dialect, LLC (“Dialect”)            |
| Second Amended Complaint | Compl.: Ex.1007                     |
| xx:yy–zz                 | column xx, lines yy to zz           |

## I. INTRODUCTION

Petitioner submits this Petition for *Inter Partes* Review of claims 7 and 8 (the “Challenged Claims”) of U.S. Patent No. 9,495,957 (the “957 Patent” (Ex.1000)), assigned to Patent Owner (“PO”). Petitioner respectfully submits that the Challenged Claims of the 957 Patent are unpatentable under 35 U.S.C. §103 in view of the prior art references discussed herein. Accordingly, it is respectfully requested that the Board institute an *inter partes* review of the 957 Patent.

## II. MANDATORY NOTICES

### A. Real Party-in-Interest

Apart from Petitioner Microsoft Corporation, another real party-in-interest for this Petition is Bank of America, N.A.

### B. Related Matters

The 957 Patent is asserted by PO in *Dialect, LLC v. Bank of America, N.A.*, 2:24-cv-00207-JRG (EDTX). In addition, the following cases involve PO and related patents: *Dialect, LLC v. Salesforce, Inc.*, 7:25-cv-00061, *Dialect, LLC v. Microsoft Corporation*, 2-24-cv-01067 (EDTX), *Google LLC v. Dialect, LLC*, IPR2024-00750 (PTAB), *Google LLC v. Dialect, LLC*, IPR2024-00752 (PTAB), *Google LLC v. Dialect, LLC*, IPR2024-00753 (PTAB).

### C. Lead and Back-Up Counsel and Service Information

| Lead Counsel                                                                                                                                                    | Back-Up Counsel                                                                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Scott M. Border</b><br>(Reg. #77,744)<br>Winston & Strawn LLP<br>1901 L Street, N.W.<br>Washington, D.C. 20036<br>Tel: (202) 282-5054<br>sborder@winston.com | <b>Carson Swope</b><br><i>(Pro Hac Vice to be submitted)</i><br>Winston & Strawn LLP<br>255 Shoreline Dr., Ste. 520<br>Redwood City, CA 94065<br>Tel: (650) 858-6407<br>cswope@winston.com |
|                                                                                                                                                                 | <b>T.D. Goswami</b><br>(Reg. #78,054)<br>Winston & Strawn LLP<br>1901 L Street, N.W.<br>Washington, D.C. 20036<br>Tel: (202) 282-5309<br>tdgoswami@winston.com                             |

Petitioner consents to electronic service by email at Winston-Microsoft-Dialect-IPR@winston.com and the e-mail addresses listed above.

### III. PAYMENT OF FEES

Petitioner authorizes the Office to charge the filing fee and any other necessary fee to Deposit Account No. 501814.

### IV. REQUIREMENTS FOR *INTER PARTES* REVIEW

#### A. Grounds for Standing

Petitioner certifies that the 957 Patent is available for inter partes review. Petitioner is not barred or estopped from requesting an inter partes review challenging the claims on the identified grounds herein. Petitioner has not filed a civil action challenging the validity of a claim of the 957 Patent. This petition is

being filed no more than 1 year after the date on which Petitioner was served with a complaint alleging infringement of the 957 Patent.

**B. Identification of Challenged Claims**

**Ground 1:** Maes (Ex. 1005) renders obvious claims 7 and 8 under 35 U.S.C. § 103.

**Ground 2:** Maes and Coffman (Ex. 1012) render obvious claims 7 and 8 under 35 U.S.C. § 103.

**Ground 3:** Kennewick (Ex.1032), Weissman (Ex.1025) and Maes render obvious claims 7 and 8 under 35 U.S.C. § 103.

**V. THE 957 PATENT**

**A. Effective Filing Date**

Petitioner assumes for the purposes of this Petition that August 29, 2005 is the effective filing date.

**B. Person of Ordinary Skill in the Art**

A POSITA with respect to the Asserted Patents as of the time of the invention, i.e., the earliest possible priority date of the 957 Patent, August 29, 2005, would have had a bachelor's degree in electrical engineering, computer science, computer engineering, or an equivalent, and two years of relevant experience involving computer science fundamentals, including natural language processing, speech recognition and transcription, non-speech recognition and transcription that is

pertinent to the 957 Patent. Lack of professional experience can be remedied by additional education, and vice versa. Ex.1003, ¶25.

### **C. Overview of the 957 Patent**

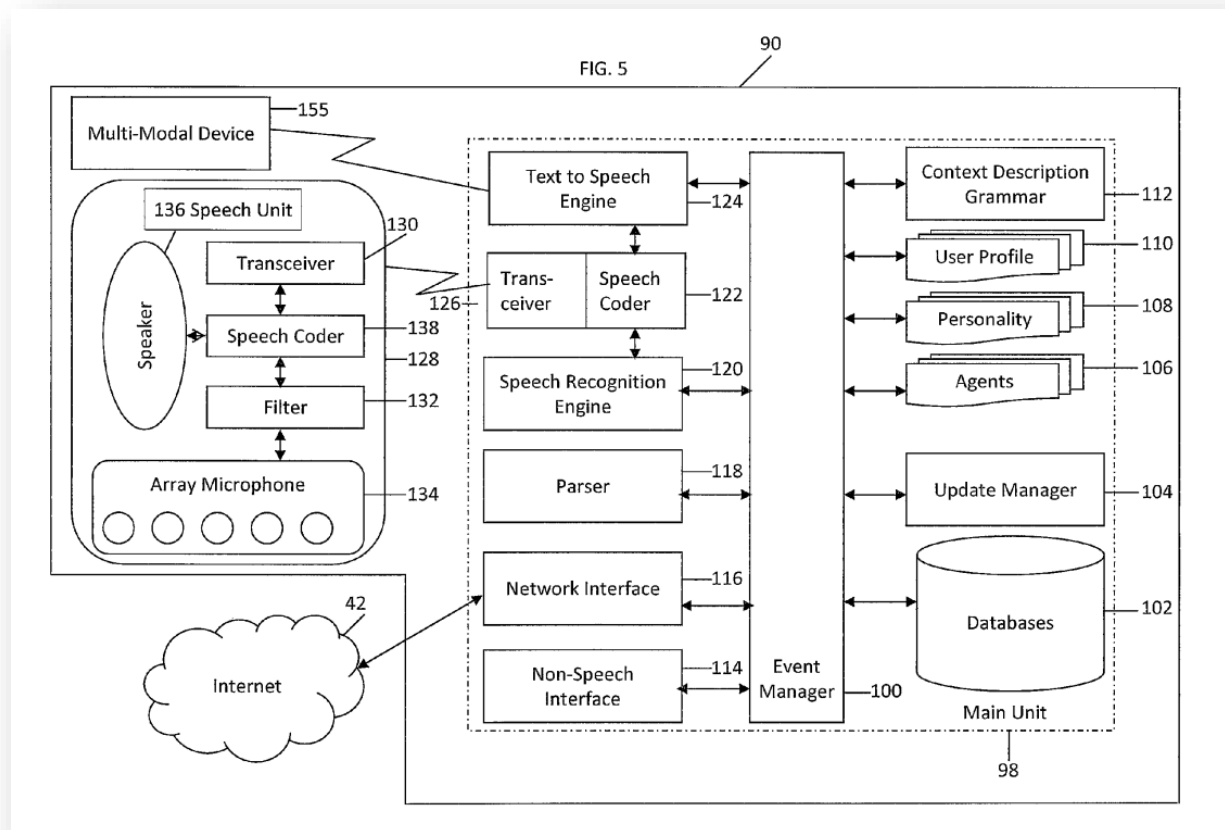
The 957 Patent was based on U.S. Patent Appl. No. 14/467,641 filed on August 25, 2014, and claims earliest priority to U.S. Patent Appl. No. 11/212,693, filed on August 29, 2005 (now U.S. Patent No. 7,949,529). Ex.1001, Cover.

The 957 Patent explains that it “provides mobile devices having a speech interface and/or a combination of a speech interface and a non-speech interface to enable natural language human-machine interactions” and that “[m]ore specifically, the invention enables mobile users to submit natural language speech and/or non-speech questions or commands in a wide range of domains.” Ex.1001, 1:24–30. The 957 Patent states that it “creates, stores and uses extensive personal profile information for each user, thereby improving the reliability of determining the context and presenting the expected results for a particular question or command. The invention may organize domain specific behavior and information into agents, that are distributable or updateable over a wide area network.” Ex.1001, Abstract.

The 957 Patent notes that “[s]peech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications,” and that “[n]atural language processing has been applied to the parsing of speech queries.” Ex.1001, 2:2–6. The 957 Patent also states that there apparently “remain a number

of significant barriers to creation of a complete natural language verbal and/or textual-based query and response environment.” Ex.1001, 2:10–12. The 957 Patent purportedly overcomes the “drawbacks” of prior art systems. Ex.1001, 2:61–62.

Figure 5 of the 957 Patent is characterized as “an overall diagrammatic view of the interactive natural language processing system according to an embodiment of the invention.” Ex.1001, 15:65–67.



As depicted in Figure 5, the disclosed system comprises speech unit 128, speech recognition engine 120, context description grammar module 112, parser

118, agents 106, multi-modal device 155 and non-speech interface 114. Speech unit 128 includes a microphone to receive a spoken utterance from a user. Ex.1001, 26:27–31.

After initial processing, the speech is passed to speech recognition engine 120 for processing using the context description grammar module 112. Ex.1001, 26:64–66. After that, “[a]ny recognized information may be processed by the parser 118, which transforms information into complete algorithms and questions using data supplied by knowledge agents.” Ex.1001, 26:67–27:2. “The knowledge agents may then process the commands or questions” and “may return results” to the user. Ex.1001, 27:2–17. Moreover, “multimodal support may be provided to maintain the context during both voice interaction and interaction through the non-speech interface 114,” Ex.1001, 27:38–41, and “multi-modal devices 155 ... may be used as a substitute or complement to the speech interface.” Ex.1001, 27:33–34.s

## **VI. CLAIM CONSTRUCTION**

Claims are given their “ordinary and customary meaning” as understood by a POSITA and the prosecution history pertaining to the patent. 37 C.F.R. § 42.100(b). Because a POSITA would find the challenged claims unpatentable under any interpretation consistent with their plain and ordinary meaning in the context of the 160 Patent, the Board need not expressly construe the claim terms. See *Vivid Techs., Inc. v. Am. Sci. & Eng. ’g. Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999).

## VII. PRINCIPAL PRIOR ART

### A. Summary of Maes

U.S. Patent No. 6,964,023 to Maes et al. (Ex.1005) was filed on February 5, 2001, and issued on November 8, 2005. Therefore, Maes qualifies as prior under at least 35 U.S.C. § 102(e).

Maes describes “systems and methods are provided for performing focus detection, referential ambiguity resolution and mood classification in accordance with multi-modal input data ... to provide an effective conversational computing environment for one or more users.” Ex.1005, Abstract.

Moreover, Maes expressly incorporates by reference Coffman in its entirety. In particular, Maes states that “[i]t is to be appreciated the conversational virtual machine disclosed in PCT patent application identified as PCT/US99/22927 filed on Oct. 1, 1999 and entitled “Conversational Computing Via Conversational Virtual Machine,” [*i.e.*, Coffman] the disclosure of which is incorporated by reference herein, may be employed to provide a framework for the I/O manager, recognition engines, dialog manager and context stack of the invention.” Ex.1005, 9:30–37. Coffman is, therefore, considered effectively part of Maes. *Arbutus Biopharma Corp. v. ModernaTX, Inc.*, 65 F.4th 656, 662-63 (Fed. Cir. Apr. 11, 2023) (“When a reference or material from various documents is incorporated, they are ‘effectively part of the host document as if [they] were explicitly contained therein.’”); *see also*

*Advanced Display Sys., Inc. v. Kent State Univ.*, 212 F.3d 1272, 1282 (Fed. Cir. 2000) (“Material not explicitly contained in the single, prior art document may still be considered for purposes of anticipation if that material is incorporated by reference into the document.”); *see also Harari v. Lee*, 656 F.3d 1331, 1335 (Fed. Cir. 2011).

As the entire disclosure of Coffman is directed to “Conversational Computing Via Conversational Virtual Machine,” Dr. Houh explains that a POSITA would understand Maes to be incorporating with detailed particularity the entire disclosure of Coffman. At a minimum, the disclosures related to the “conversational virtual machine” that “provide a framework for the I/O manager, recognition engines, dialog manager and context stack,” are expressly incorporated with particularity, which is the material relied upon below. Ex.1003, ¶55.

#### **B. The Combination of Maes’s Teachings with those of Coffman**

To the extent one asserts that Coffman discloses a different embodiment from that of Maes, a POSITA would have been motivated to combine Coffman’s teachings related to searching a context stack to find entries matching transcribed user input and merging/combining speech and non-speech transcriptions (Ex.1012, 41:2–5, 42:1–3, and 38:1–19) with Maes’s teachings that receives and processes multi-modal inputs (e.g., by producing decoded text or script) as described below. Ex.1003, ¶56.

As explained above, Maes expressly incorporates by reference Coffman in its entirety, names Stephane Maes as a common inventor, and has IBM as the original

assignee. Ex.1005, Abstract, 9:30–37 and 19:43–49 and Ex.1012, Cover, Abstract. Accordingly, a POSITA would have been motivated to combine the teachings of Coffman with those of Maes for at least these reasons. Ex.1003, ¶57. *Black v. CE Soir Lingerie Co.*, No. 2:06-CV-544, 2008 WL 3852722, at \*14 (E.D. Tex. Aug. 15, 2008) aff’d, 319 F. App’x 901 (Fed. Cir. 2009). (“strong motivation to combine” prior art “patented by the same inventor” and that “reference each other.”); *Abbot Vascular, Inc. v. Flexstent*, IPR2019-00882, Paper 48, 28-29 (Oct. 2, 2020) (finding that a POSITA would be motivated to combine references by the same author and from the same company).

Combining Coffman’s teachings of organizing/updating its context stack such that the most-recent context is at the top of the context (Ex.1012, 21, 23–28, 41:1–5, 40:15–19, and 42:7–9) with the teachings of Maes would benefit Maes’s system by providing efficient identification and retrieval of the most-recent context, allowing for faster executions for matching/decoding of keywords and improving the responsiveness of Maes’s system aligning with its goal of faster executions (Ex.1005, 11:64–12:30, 5:29–34). A POSITA would have understood that by keeping the context stack updated, e.g., with the most-recent context associated with a user’s dialog, would allow an application to use the most-recent context, which is likely to be relevant to the user’s utterance, in determining the user’s intent. Ex.1005, 36:64–37:3. Ex.1003, ¶58.

A POSITA would be motivated to combine Maes and Coffman because they teach strikingly similar techniques. For example, Maes's technique of associating recognized events against data in a context stack (Ex.1005, 7:60–8:4 and Figures 1 and 8), and Coffman's technique of searching a context stack to find entries matching transcribed user input (Ex.1012, 42:1-9). Notably, Maes's context stack disclosures are almost verbatim as Coffman's. Ex.1012, 21, 23–28 and Ex.1005, 37:51–62; *compare also* Ex.1005, Figure 8B and Ex.1012, Figure 2. Ex.1003, ¶59.

Additionally, a POSITA would have understood that combining Maes's teachings and Coffman involves the application of known techniques (e.g., identifying contexts by performing comparisons of spoken utterances against data stored in a context stack) to improve a similar system (e.g., Maes's system) in the same way. Ex.1003, ¶60.

Maes and Coffman are analogous art as the 957 Patent for at least three reasons. *First*, both references are directed toward the same field of endeavor as the 957 Patent—computer-implemented systems interpreting user utterances or otherwise natural language processing systems. Ex.1001, Abstract, 1:51–61, 3:25–45; Ex.1005, Abstract, 4:7–17; Ex.1012, 1–4 and 29:1–4. *Second*, Maes and Coffman are reasonably pertinent to a problem described by the 957 Patent: an environment for **reliably processing** a user's language queries, which is a problem of conventional systems identified in the 160 Patent. Ex.1001, 1:35-44; Ex.1005,

Abstract, 4:7–17 and 6:39–55; Ex.1012, 4:4–9. *Third*, similar to the 957 Patent, Maes and Coffman teach the use of **grammars** and **vocabularies** in recognizing speech inputs, which indicate compatibility of their teachings and therefore a POSITA would have had a reasonable expectation of success combining them. Ex.1001, 13:35–38; Ex.1005, 31:15–18, 34:61–67, 33:11–21, 39:59, 41:13–20, and 41:35–38; Ex.1012, 29:21–32, 25:6–9. Ex.1003, ¶61.

A POSITA would be motivated to combine Maes and Coffman because to do so would have been the arrangement of old elements (a speech transcription and recognition system comprising modules for speech transcription and recognition, context stack storing current and historical data related to user interactions) with each performing the same function it has been known to perform (e.g., recognizing spoken utterances, converting spoken utterances into computer-readable format based on decoding the utterances, searching a context stack to find matches between data stored on the context stack and transcribed user input) and yielding no more than what one would expect from such an arrangement (an improved computer-implemented system interpreting user utterances), as Maes seeks to implement. Ex.1005, 11:62–63, 2:54–67. Ex.1003, ¶62.

Furthermore, a POSITA would have been motivated to combine Maes and Coffman because the references make clear that their systems require no specialized hardware or software. In fact, these references teach using conventional,

commercially-available systems. *See* Ex.1005, 2:38–53 and 45:50–60, 12:43–49 (“processing ... may be accomplished via any **conventional acoustic information recognition system**”) (emphasis added<sup>1</sup>); Ex.1012, Abstract (“The conversational computing system may be built on top of a **conventional operating system and API’s ... and conventional device hardware.**”). Ex.1003, ¶63.

A POSITA would have had a reasonable expectation of success in combining the teachings of these references because to do so would be a simple combination of analogous teachings that would be well within the skill of a POSITA. For example, modifying Maes’s system producing decoded text to include Coffman’s teachings related to searching a context stack to find entries that match transcribed user input and merging/combining speech and non-speech transcriptions would enhance the functionality of Maes’s system and further, as explained below, such modifications would have been simple implementations predictably resulting in a recognition system with improved accuracy and without undue experimentation, as Maes seeks to implement (Ex.1005, 11:62–63). Ex.1003, ¶64.

---

<sup>1</sup> Emphasis added throughout unless otherwise specified.

## VIII. THE CHALLENGED CLAIMS ARE UNPATENTABLE

### A. Ground 1: Maes renders obvious claims 7 and 8

#### 1. Claim 7

- a. [7.0] A computer-implemented method of processing a natural language utterance, the method comprising:

Maes describes “[s]ystems and methods ... provided for performing focus detection, referential ambiguity resolution and mood classification **in accordance with multi-modal input data [in the form of audio input data, video input data, as well as other types of input data].**” Ex.1005, Abstract (emphasis added<sup>2</sup>), 4:10–11, 1:9–23; *see also* 1:9–23; Ex.1003, ¶66.

Further, Maes’s systems and methods are directed at “*processing a natural language utterance.*” While Maes discusses processing multi-modal inputs, the multi-modal input may comprise, at least in part, “audio input data” (*e.g.*, “*a natural language utterance*”). Referencing Figure 1 (below), Maes discloses:

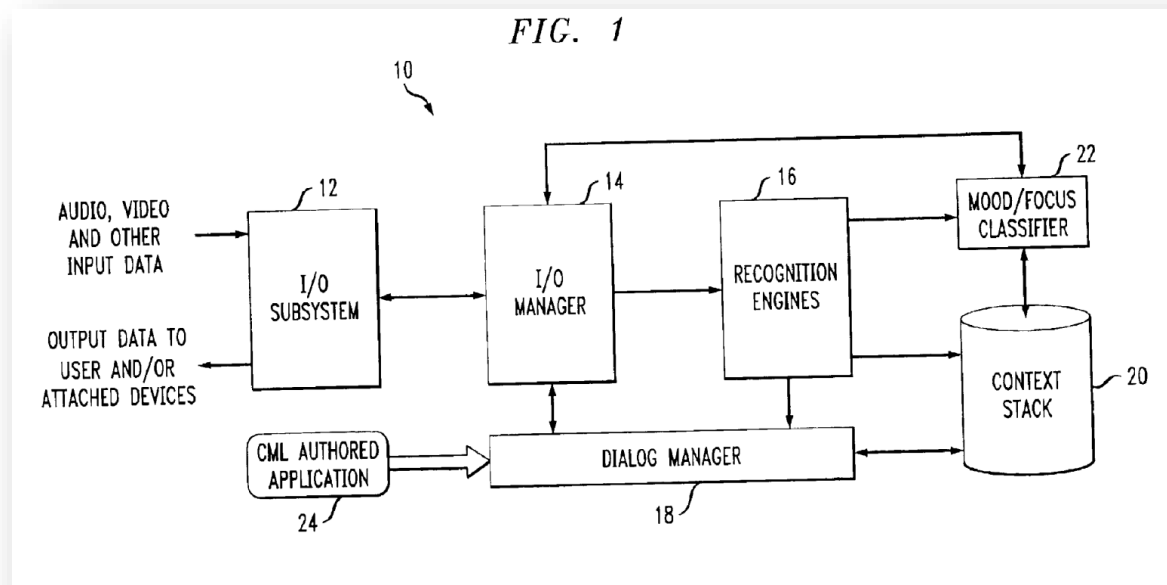
[A] multi-modal conversational computing system ... compris[ing] an input/output (I/O) subsystem 12, an I/O manager module 14, one or more recognition engines 16, a dialog manager module 18, a context stack 20 and a mood/focus classifier 22.

---

<sup>2</sup> Emphasis added throughout unless otherwise specified.

**[T]he multi-modal conversational computing system 10 of the present invention receives multi-modal input in the form of audio input data, video input data, as well as other types of input data (in accordance with the I/O subsystem 12), processes the multi-modal data (in accordance with the I/O manager 14), and performs various recognition tasks (e.g., speech recognition ... in accordance with the recognition engines 16), if necessary, using this processed data. The results of the recognition tasks and/or the processed data, itself, is then used to perform one or more conversational computing tasks, e.g., focus detection, referential ambiguity resolution, and mood classification (in accordance with the dialog manager 18, the context stack 20 and/or the classifier 22).**

*See generally* Ex.1005, 3:66–4:21.



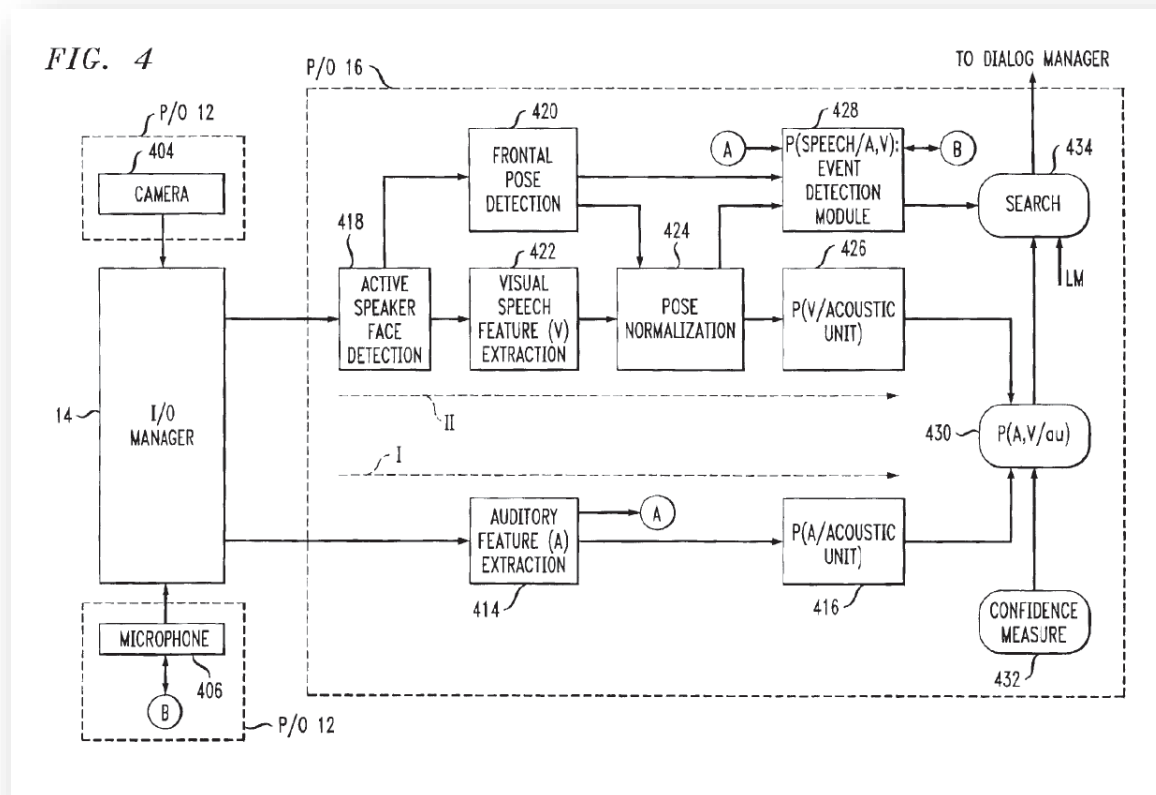
Ex.1005, Figure 1; Ex.1003, ¶68.

Maes describes Figure 4 as “an audio-visual speech recognition module<sup>3</sup> that may be employed as one of the recognition modules<sup>4</sup> of FIG. 1 **to perform speech recognition using multi-modal input data received** in accordance with the invention.” Ex.1005, 10:50–52. Maes notes that the “microphone 406” of Figure 4 is “part of the I/O subsystem 12” (Ex.1005, 11:24–26) which “comprise[s] one or more microphones for capturing audio input data from the environment in which the system is deployed.” Ex.1005, 6:5–8; Ex.1003, ¶69.

<sup>3</sup> As used herein, the terms “Maes’s system” and “audio-visual speech recognition module” are used interchangeably.

<sup>4</sup> Maes uses the term “block” and “module” interchangeably. For example, see 12:43-38 and 17:46-50 referring to “414” as “block[] 414” and “module 414.”

Figure 4 shows that the audio-visual speech recognition module receives multi-modal input data—notably, including an audio signal (e.g., speech provided by the speaker and/or background noise) (“*a natural language utterance*”) and a video signal (e.g., the speaker’s face including lip movement and/or background objects in the environment). Video signals are processed by blocks/modules 418, 422, 424, 426 and audio signals are processed by blocks/modules 414, 416.



Ex.1005, Figure 4; Ex.1003, ¶70.

Maes’s methods are “*computer implemented method[s]*.” For example, Maes discloses that, “**computer software including instructions or code for performing**

**the methodologies of the invention** ... may be stored in one or more of the associated memory devices (e.g., ROM, fixed or removable memory) and, when ready to be utilized, loaded in part or in whole (e.g., into RAM) and executed by a CPU. ... [T]he elements illustrated in FIGS. 1 through 9C may be implemented in various forms of hardware, software, or combinations thereof, e.g., one or more digital signal processors with associated memory, application specific integrated circuit(s), functional circuitry, one or more appropriately programmed general purpose digital computers with associated memory.” Ex.1005, 45:45–60; Ex.1003, ¶71.

Maes describes examples in which its methods are applicable to a broadcast news system which uses spoken utterance<sup>5</sup> inputs which further confirms that its method applies to “*processing a natural language utterance.*” Maes’s system provides, among others, functions of “natural language understanding (NLU), natural language generation (NLG)” and “NL parsing” can be used in the context of broadcast news which contain a newscaster speaking at a location where there is arbitrary activity and noise in the background (Ex.1005, 41:14–16, 35:58, 10:63–11:15) and may be employed within a vehicle such that if a user says the spoken

---

<sup>5</sup> As used herein, it will be understood that a “spoken utterance” received using Maes’s system, would encompass both an audio portion and a video (i.e., non-speech) portion (e.g., the accompanying lip movement) of the utterance. Ex.1005, 6:43–47.

utterance “turn it on,” Maes’s system processes inputs relating to an I/O event representative of the user’s spoken utterance. Ex.1005, 4:30–32, 6:43–50, and 7:63–8:29; Ex.1003, ¶¶72–73.

Referencing the in-vehicle system example, Maes explains a user in the vehicle may say “turn it on,” while pointing at the vehicle radio. “The I/O manager [of the system] receives the raw multi-modal data and abstracts the data into a form that represents ... a spoken utterance .... As is known, a data abstraction operation may involve generalizing details associated with all or portions of the input data ... to yield a more generalized representation of the data for use in further operations.” Ex.1005, 6:35–38. “The dialog manager would therefore receive the results of the recognized events associated with the spoken utterance ‘turn it on’” including the accompanying lip movement. Ex.1005, 7:63–67; *see also* 6:43–50; Ex.1003, ¶74.

Therefore, if limiting, Maes discloses the preamble. Ex.1003, ¶75.

- b. **[7.1] generating a context stack comprising context information that corresponds to a plurality of prior utterances, wherein the context stack includes a plurality of context entries;**

Maes describes “the context stack 817 may be **implemented ...**” (“*generating a context stack*”) as part of the context stack 20 of Figure 1. Ex.1005, 37:53–55. Referencing Figure 1 (below), Maes describes “the multi-modal conversational

computing system 10 comprises ... a context stack 20.” Ex.1005, 3:66–4:6; Ex.1003, ¶77.

Maes further describes its “*context stack*” as storing “historical information” e.g., past input/output (I/O) events or “*a plurality of context entries*” generated previously in the context of a dialog (e.g., “*context information that corresponds to a plurality of prior utterances*”).

[H]istorical information (e.g., past events) stored in the context stack [such that] the context stack ... is associated with the **organized/sorted context** corresponding to each active **dialog**.

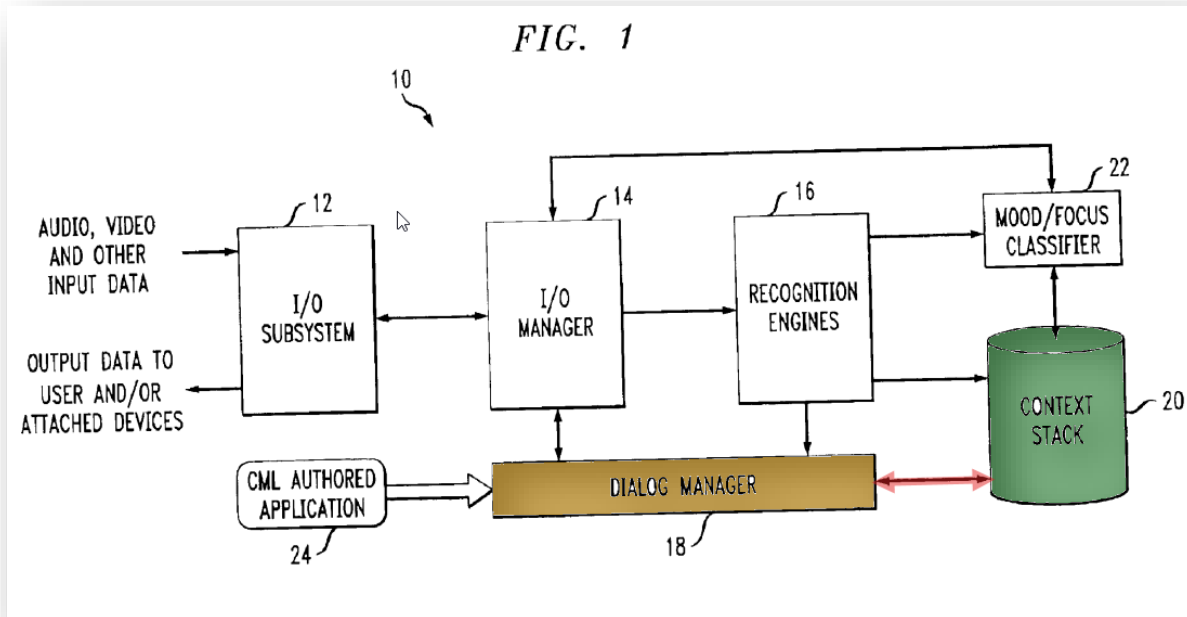
Ex.1005, 7:62–63, 37:60–61; Ex.1003, ¶78.

Maes’s context stack “*compris[es] context information*” and “*includes a plurality of context entries*” because Maes teaches

**context stack 817 comprises** all the information associated with an application ... [which] includes **all the variable, states, input, output and queries to the backend that are performed in the context of the dialog** and any extraneous event that occurs during the dialog. The context stack is associated with the organized/sorted context corresponding to each active dialog.

Ex.1005, 37:55–62; Figure 1 (reproduced below); *see also* Ex.1005, 7:40–45 (disclosing “the system ... attempts to determine the **user intent based on the**

current event and the historical interaction information stored in the context stack.”); 5:16–19 (disclosing “an I/O (input/output) event generated previously and stored in a context manager/history stack (e.g., if a beeper rang and the user asked ‘turn it off’)”).



Ex.1005, Fig. 1 (annotated); Ex.1003, ¶79.

Referencing the in-vehicle example in which a user may say “turn it on,” Maes describes its system identifying words in a recognized utterance from contexts—such as by matching the utterance “turn it on” to the context “radio”:

The dialog manager [such as dialog manager 18 in Figure 1] would ... receive the results of the recognized events associated with the spoken utterance “turn it on” and the gesture of pointing to the radio. Based on these events, the

dialog manager does a search of the **existing applications, transactions or ‘dialogs,’** [*“context information that corresponds to a plurality of prior utterances a plurality of prior utterances”*] or portions thereof [stored on the context stack], with which such an utterance and gesture could be associated.

Ex.1005, 7:65–8:4

**[T]his recognized spoken utterance event is stored on the context stack.** Then, when the recognized gesture event (e.g., pointing to the radio) is received, the dialog manager takes this event and **the previous spoken utterance event stored on the context stack and makes a determination that the user intended to have the radio turned on.**

Ex.1005, 8:37–42; Ex.1003, ¶80.

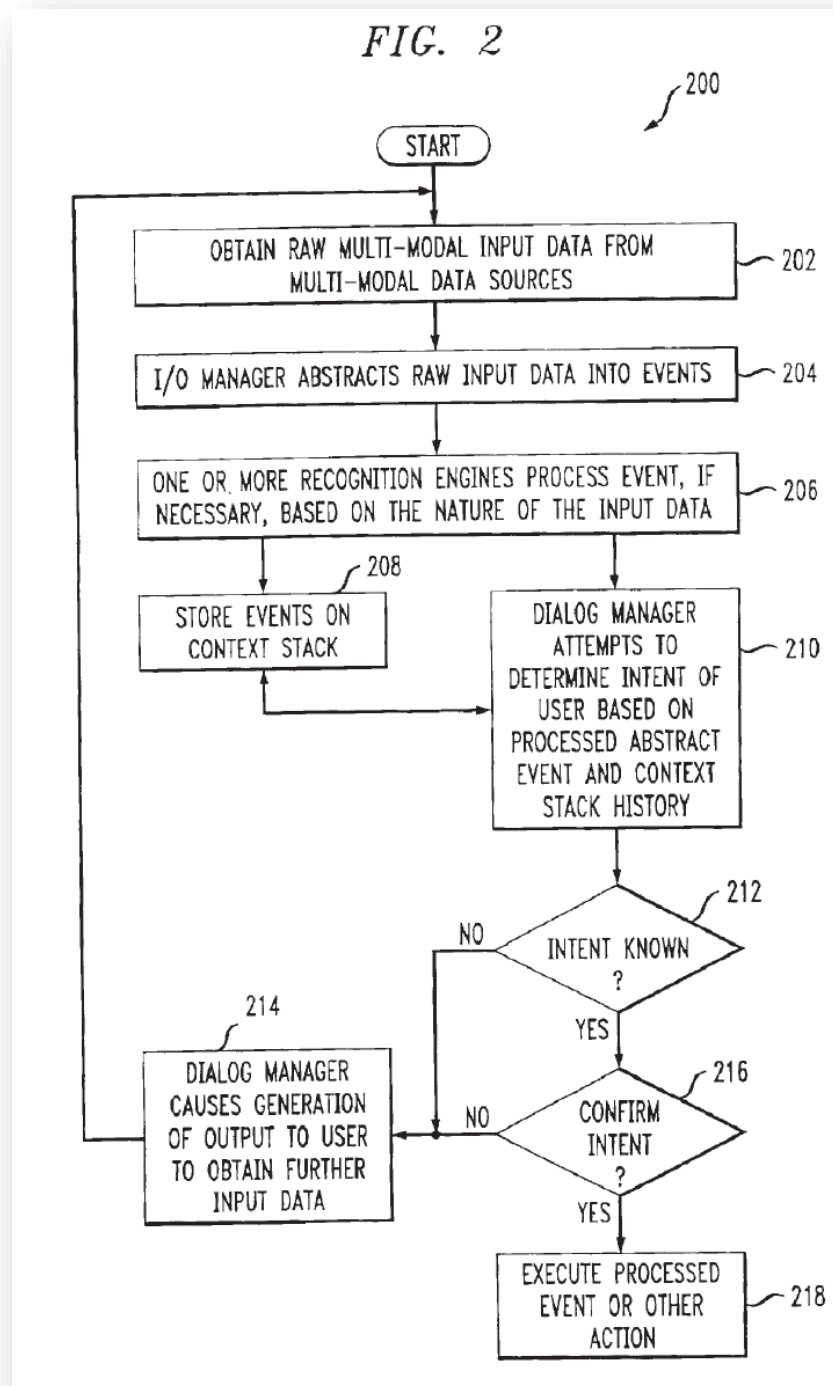
In the above example, “this event” corresponds to the current event of the user pointing to the radio, which occurred after “the previous spoken utterance event” of the user saying “turn it on.” Data for both events (*“context information”*) are stored on the context stack because Maes discloses **“all the variable, states, input, output and queries to the backend that are performed in the context of the dialog and any extraneous event that occurs during the dialog”** are stored on the context stack. Ex.1005, 37:55–62. Maes clarifies that “the dialog” includes

“**conversational dialog comprising speech** and other multi-modal I/O such as GUI keyboard, pointer, mouse, as well as video input, etc.” Ex.1005, 36:49–54; Ex.1003, ¶80.

Therefore, Maes teaches or suggests this claim element at least for this reason. Ex.1003, ¶81.

### **Additional Mapping**

In addition, Figure 2 described as “a flow diagram illustrating a referential ambiguity resolution methodology performed by a multi-modal conversational computing system” expressly discloses the use of context stack for storing event-related data and used in determining the user’s intent. Ex.1005, 3:24–28.



Ex.1005, Figure 2; Ex.1003, ¶82.

Referencing Figure 2, Maes explains that in step 208, “**stor[ing]** [“*generating*”] **[recognized events, as well as the events that do not need to be recognized]** in a storage unit referred to as the context stack 20 ... used to create a history of interaction between the user and the system [“*a context stack comprising context information that corresponds to a plurality of prior utterances, wherein the context stack includes a plurality of context entries*”] ... to assist the dialog manager ... in making referential ambiguity resolution determinations when determining the user’s intent. Next, in step 210, the system ... attempts to determine the user intent based on the current event and the **historical interaction information stored in the context stack.**” Ex.1005, 7:33–45; Ex.1003, ¶83.

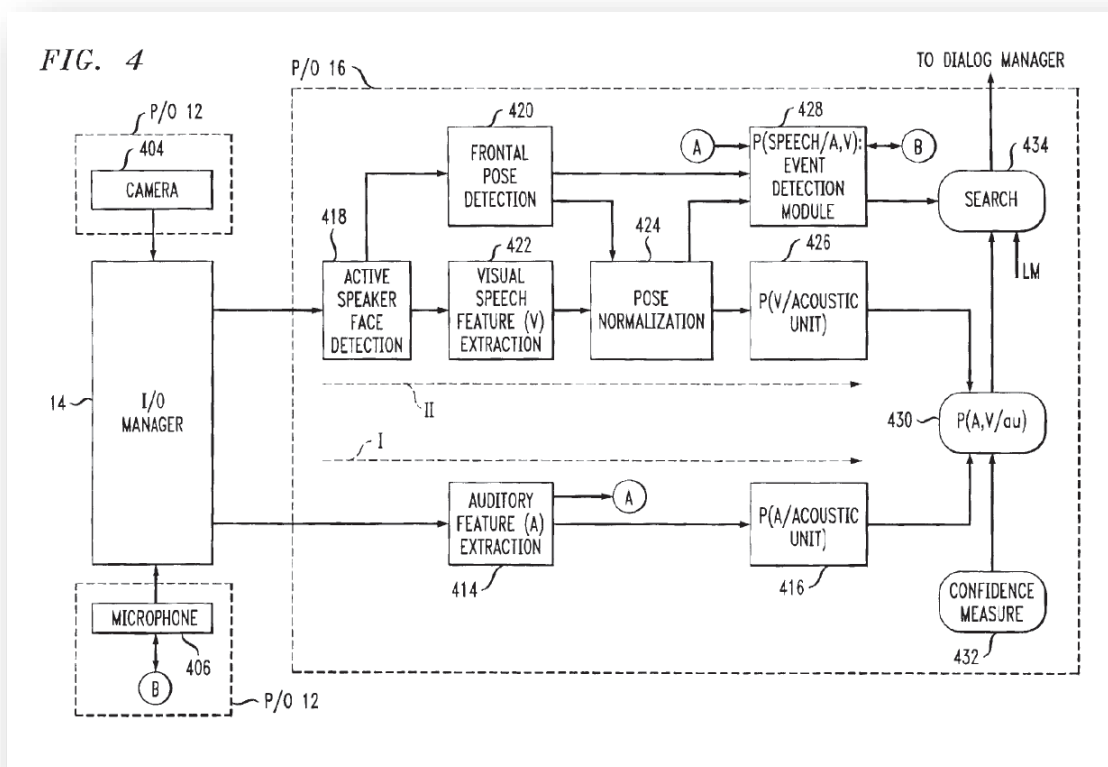
- c. **Therefore, Maes teaches or suggests this claim element for this additional reason. Ex.1003, ¶84. [7.2] receiving, by one or more computer processors, the natural language utterance, wherein the natural language utterance is associated with a command or is associated with a request;**

- i. **Receiving the natural language utterance**

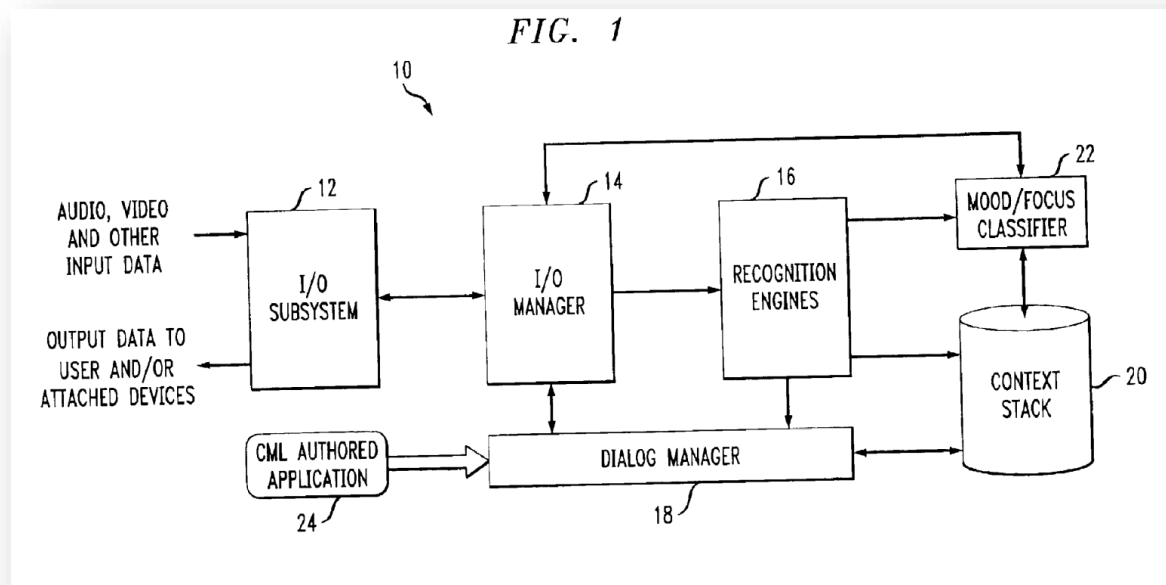
For example, referencing Figure 4, Maes discloses “an audio-visual speech recognition module,” “receive[s]” (“*receiving*”) multi-modal input data—such as audio and visual data—from a speaker (e.g., comprising “*the natural language utterance*”) at an “input/output (I/O) subsystem ... compris[ing] one or more **microphones for capturing audio input data,**” which “one of ordinary skill in the art will realize that other user interfaces and devices ... may be included for

capturing user activity.” Ex.1005, 10:50–52, 6:21–29, 5–8; *see also* 11:24–26. Ex.1003, ¶86.

As Dr. Houh explains, Maes describes its system includes functionality for capturing audio signal, video signal, or both, as part of receiving input data associated with the natural language utterance.



Ex.1005, Figure 4.



Ex.1005, Figure 1; Ex.1003, ¶86 (citing 5:60–64).

**ii. one or more computer processors**

Maes's “**system ... comprises at least one processor**, operatively coupled to the user interface subsystem ... configured to receive ... **the multi-modal input data** from the user interface subsystem [and] configured to ... make a determination of at least one of an intent, a focus and a mood of at least one of the one or more users based on at least a portion of the received multi-modal input data. The processor is ... further configured to ... cause execution of one or more actions to occur in the environment based on at least one of the determined intent, the determined focus and the determined mood. The system further comprises a memory, operatively coupled to the **at least one processor**, which stores at least a

portion of results associated with the intent, focus and mood determinations made by the processor for possible use in a subsequent determination or action.” Ex.1005, 2:38–54; Ex.1003, ¶87.

Maes explains that “[T]he elements illustrated in FIGS. 1 through 9C may be implemented in various forms of hardware, software, or combinations thereof, e.g., one or more digital signal processors with associated memory, application specific integrated circuit(s), functional circuitry, one or more appropriately programmed general purpose digital computers with associated memory, etc. Given the teachings of the invention provided herein, one of ordinary skill in the related art will be able to contemplate other implementations of the elements of the invention.” Ex.1005, 45:50–60; Ex.1003, ¶88.

**iii. The natural language utterance is associated with a command/request**

Maes provides an example of an in-vehicle system in which a user may say the “spoken utterance ‘turn it on’” (which Maes alternatively refers as “the ‘turn it on’ command”) (“*the natural language utterance ... associated with a command or ... associated with a request*”). Ex.1005, 8:30–33. Maes’s system processes inputs relating to an I/O event representative of the user’s spoken utterance (audio portion) and the accompanying lip movement of the spoken utterance (video portion). Ex.1005, 4:30–32, *see generally* 7:63–8:29; Ex.1003, ¶89.

Maes explains that “the microphone picks up the audible portion of the utterance and a camera picks up the visual portion (e.g., lip movement) of **the utterance** [*“the natural language utterance”*], [and] the event may be sent to an audio-visual speech recognition engine to have **the utterance** recognized using ... the audio input and the video input associated with the speech.” Ex.1005, 6:43-50; Ex.1003, ¶90.

Maes also discloses that its system:

provides the capability to ... **understand queries based on who said or did what, what was the focus of the user when he gave a multi-modal query/command** [*“the natural language utterance ... associated with a command or ... associated with a request”*] ... .

Ex.1005, 2:54–67. Maes’s “system ... recognizes **voice-based commands** [*“command[s]”*] and ... gesture-based commands.” Ex. 1005, 2:1–4; *see also* Ex.1005, 4:58–59 (disclosing “user input **commands** received and processed by the system”); Ex.1003, ¶91.

- d. [7.3] **determining one or more words of the natural language utterance by performing speech recognition on the natural language utterance;**
  - i. **Performing speech recognition on the natural language utterance**

Maes's system “**performs various recognition tasks (e.g., speech recognition ... in accordance with the recognition engines 16).**” Ex.1005, 4:7–14; see also 34:61–35:1 (one or more recognition engines of FIG. 1). “[D]epending on the nature of the [I/O (input/output)] event, one or more recognition engines may be used to recognize the event. For example, if the event is some form of spoken utterance ..., the event may be sent to an audio-visual **speech recognition engine to have the utterance recognized** [*“performing speech recognition on the natural language utterance”*].” Ex.1005, 6:42–50; Ex.1003, ¶94.

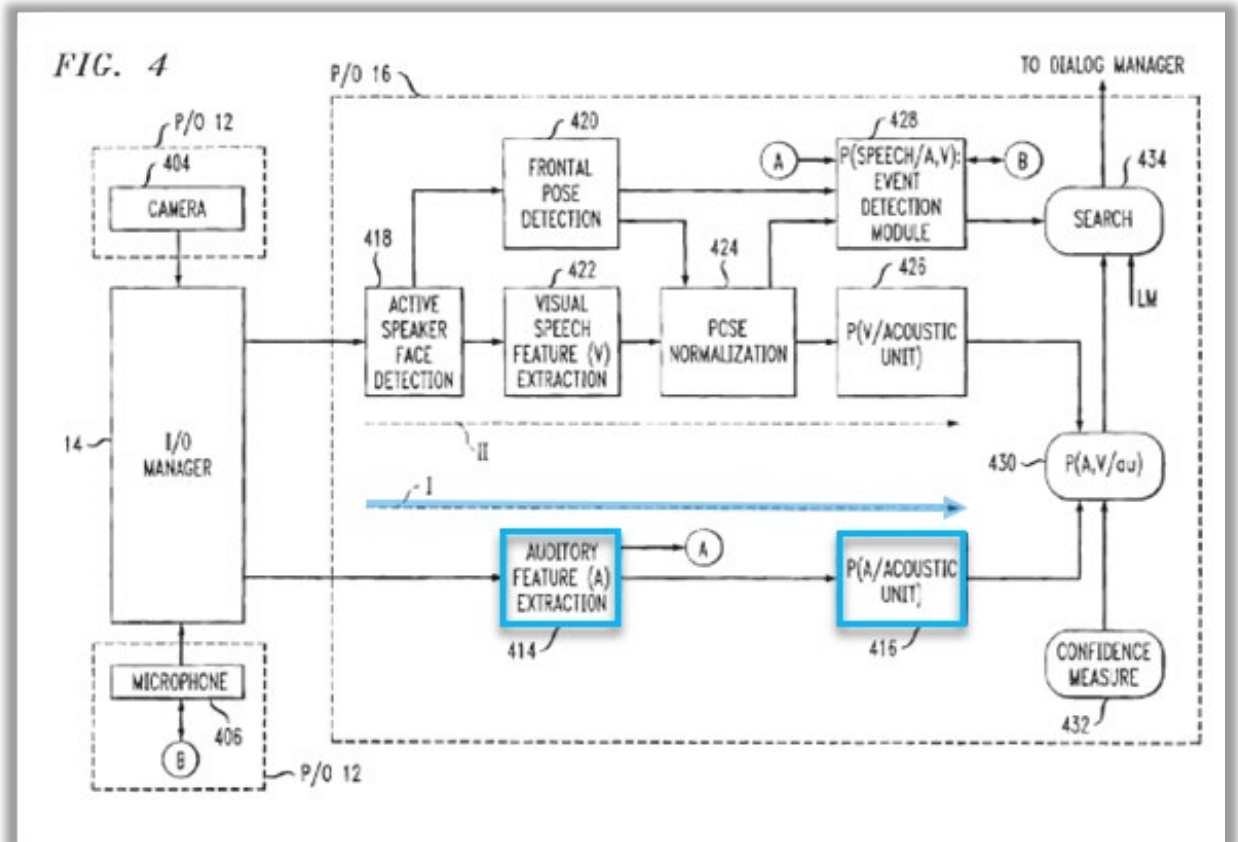
Additional details of Maes disclosing its system “*performing speech recognition engine*” are provided. Figure 4 is described as “an audio-visual **speech recognition module** that may be employed as **one of the recognition modules<sup>6</sup> of FIG. 1 to perform speech recognition using [received] multi-modal input data.**” Ex.1005, 10:50–52. Video signals are processed by blocks/modules 418, 422, 424, 426 and audio signals are processed by blocks/modules 414, 416, in Figure 4. Ex.1003, ¶95.

Maes states that “the processing performed in blocks 414 and 416 may be **accomplished via any conventional acoustic information recognition system** [*“performing speech recognition”*] capable of extracting and labeling acoustic

---

<sup>6</sup> Maes uses the term “block” and “module” interchangeably. See 12:43-38, 17:46-50.

feature vectors.” Ex.1005, 12:43–49. Dr. Houh explains that one such “acoustic information recognition system” that provides additional details on how speech is processed in Maes is described with reference to Figure 9A, explained below. Therefore, at least “blocks 414 and 416” disclose a “*performing speech recognition.*”



Ex.1005, Figure 4 (showing audio information processing path (Path I) annotated in blue); Ex.1003, ¶96.

Referencing Figures 1 and 9A, Maes explains that “via the I/O manager 14 of FIG. 1” “user-provided input data events are ... provided to” “apparatus 900”—an “apparatus for collecting data associated with a voice of a user,” including “a dialog

management unit 902 for conduct[ing] a conversation with the user,” that provides functionality including natural language understanding (NLU), natural language generation (NLG), finite state grammar (FSG), and/or text-to-speech syntheses (TTS) for machine-prompting the user....” Ex.1005, 39:22–29, 41:13–20, Figure 9A; Ex.1003, ¶97.

Maes continues: “Apparatus 900 ... includes a processing module 910” further including “**a speech recognizer 926** which ... include[s] **a speech recognition module 928**,” and “a speech prototype, language model and grammar database”. Ex.1005, 39:59, 41:35–38. Apparatus 900 can further include **a post processor 938 ... configured to transcribe user utterances and ... perform keyword spotting thereon.** ... Post processor 938 can employ speech recognizer 926.” Ex.1005, 41:48–59. Therefore, Maes discloses “*performing speech recognition on the natural language utterance.*” Ex.1003, ¶98.

**ii. Determining one or more words of the natural language utterance**

Referencing Figure 4, Maes explains “**produc[ing]**” (“*determining*”), for instance, “**decoded text**” (“*one or more words*”) “by” “the processing performed in blocks 414 and 416”—(e.g., “*performing speech recognition*”) on “the uttered speech” (“*the natural language utterance*”). Ex.1005, 21:43–47; Ex.1003, ¶99.

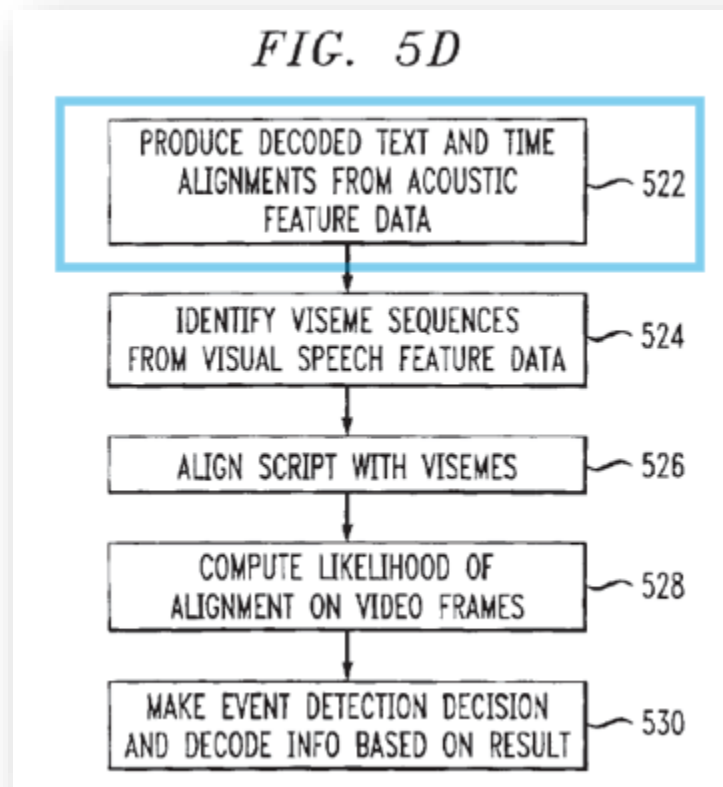
And, referencing Figure 9, Maes explains that post processor 938 includes a semantic module “to interpret meanings of phrases” and “could be used by speech recognizer 926 indicate that some **decoding candidates** in a list are meaningless and should be discarded/replaced with meaningful candidates” Ex.1005, 41:58–64; Ex.1003, ¶99.

Maes, therefore, explains that the “**decoded text**” produced by the processing performed in blocks 414 and 416 is the same “decoding candidates” Maes describes in connection with Figure 4, further confirming Figure 9A and its corresponding disclosure provide additional details to, among other things, the “*performing speech recognition*” step described in Figure 4. For example, in Figure 9A, Maes identifies “semantic module,” “speech recognizer 926,” “post processor 938,” “speech recognition module 928” and other components/functionalities. *See e.g.*, Ex.1005, 41:48–59 (**post processor 938 ... transcribe[s] user utterances and ... perform[s] keyword spotting thereon ....**”); Ex.1003, ¶99.

Therefore, Maes discloses the “*determining one or more words of the natural language utterance by performing speech recognition on the natural language utterance.*” Ex.1003, ¶100.

Referencing Figure 5D, Maes discloses that “in step 522, the uttered speech to be verified may be **decoded** by classical speech recognition techniques so that a decoded script and associated time alignments are available. This is accomplished

using the feature data from the acoustic feature extractor 414.” Ex.1005, 21:43–47. Figure 5D (block 522) discloses that “**decoded text**” (or, script) (“*one or more words*”) “produce[d]” (“*determin[ed]*”) “using the feature data from the acoustic feature extractor 414” (“*by performing speech recognition*”) “on” “the uttered speech” (“*the natural language utterance*”). *Id.*



Ex.1005, Figure 5D (block 522). Dr. Houh states that a POSITA would have understood that the “decoded script” described in Maes is the tangible result of the speech recognition system’s interpretation of spoken language and provides a text representation of the multi-modal content. Ex.1005, 20:61-64 (“If the buffered data

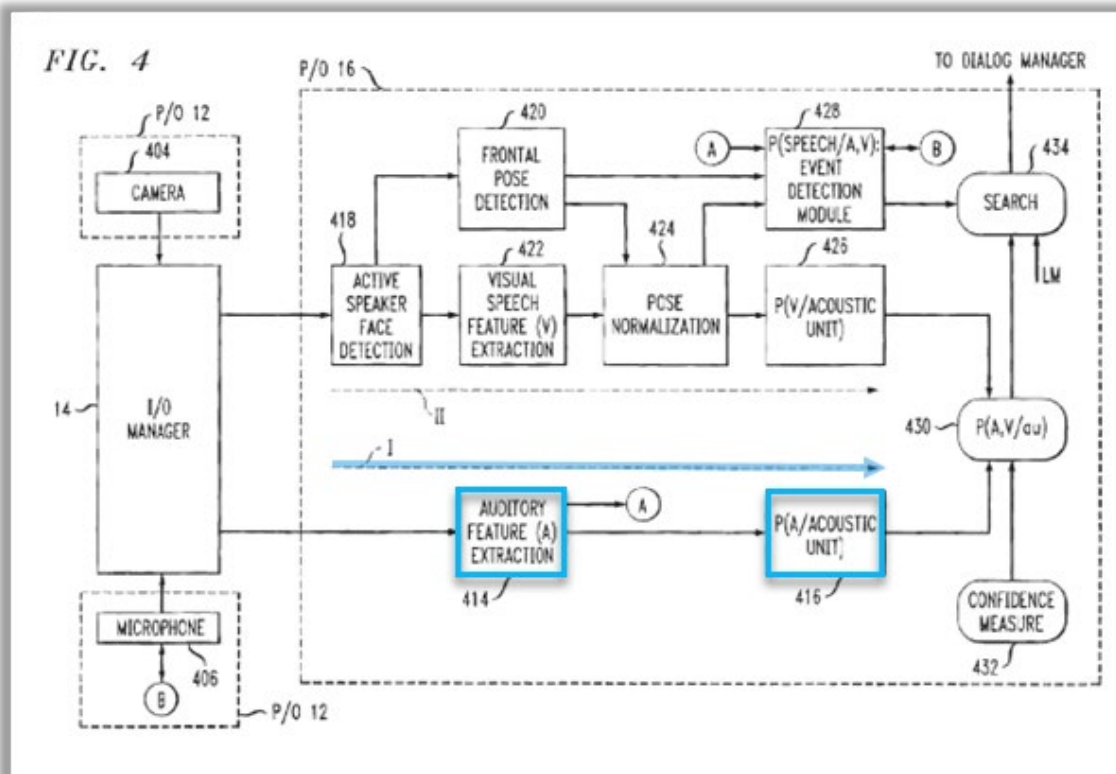
is tagged as speech, in step 518, the buffered data is sent on through the acoustic path so that the buffered data may be recognized, in step 520 ... to yield a **decoded output.**”); *see also* 20:23-33; Ex.1003, ¶101.

Referencing Figures 1 and 4, Maes provides additional details explaining the “*determining ...*” step: Maes describes Figure 4 as “an audio-visual speech recognition module that may be employed as **one of the recognition modules<sup>7</sup>** of **FIG. 1 to perform speech recognition using multi-modal input data received** in accordance with the invention.” Ex.1005, 10:50–52; Ex.1003, ¶102.

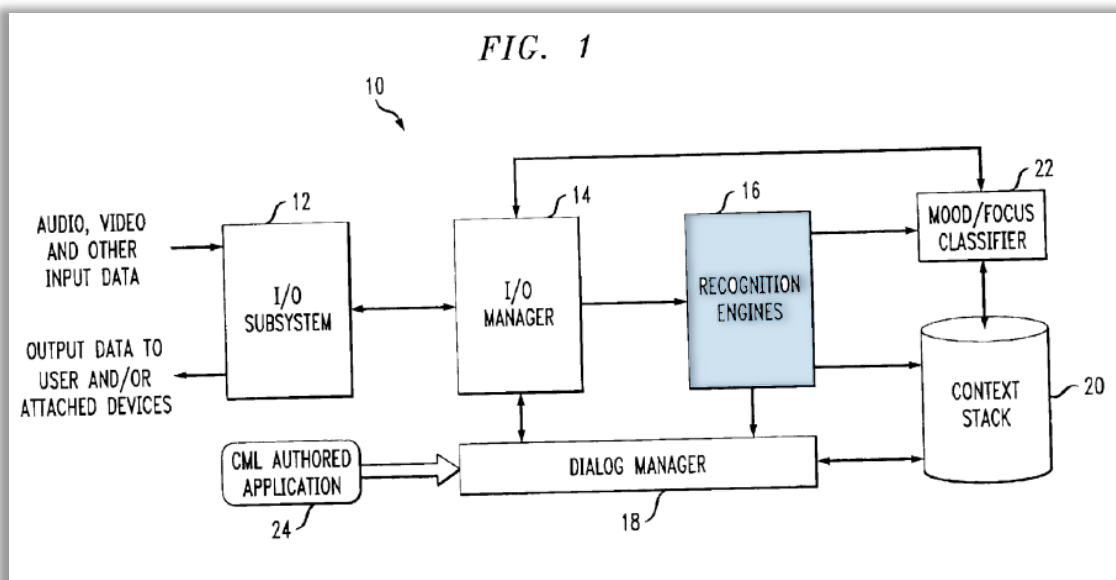
In Figure 4, the auditory feature extractor 414 “receives an audio or speech signal and, ... extracts spectral features from the signal at regular intervals[,] ... in the form of acoustic feature vectors (signals) ....” Ex.1005, 11:64–12:2.

---

<sup>7</sup> Maes uses the term “block” and “module” interchangeably.



Ex.1005, 11:55–64 and Figure 4 (showing audio information processing path (Path I) annotated in blue and acoustic feature vectors denoted by the letter “A”).



Ex.1005, Figure 1 (annotated); Ex.1003, ¶103.

Maes explains in detail how acoustic features are extracted. Ex.1005, 12:12–28; Ex.1003, ¶104. Dr. Houh explains Maes also describes sampling the signal prior to extracting the acoustic feature vectors: “Before acoustic vectors are extracted, the speech signal may be sampled at a rate of 16 kilohertz (kHz).” Ex.1003, ¶105 (citing Ex.1005, 12:2–12).

Maes provides several examples of acoustic feature vectors (or alternatively cepstral vectors) that may be extracted: e.g., [linear prediction coefficients] LPC cepstra, [Perceptual Linear Prediction] PLP, MEL cepstra, and that the invention is not limited to any particular type. Ex.1005, 23:40–43, 39:55–58; Ex.1003, ¶106.

Referencing the in-vehicle example, Maes explains the “*determining ...*” step: Maes’s system receives the raw multi-modal data relating to an I/O event and **“abstracts the data into a form that represents ... a spoken utterance ... [by] generalizing details associated with all or portions of the input data ... to yield a more generalized representation of the data for use in further operations.”** Ex.1005, 6:21–38. Maes’s system produces decoded text/script corresponding to the user’s spoken utterance, for example, of saying “turn it on.” Ex.1003, ¶107.

- e. **[7.4.0] identifying, from among the plurality of context entries, one or more context entries that correspond to the one or more words, wherein the context information includes the one or more context entries,**
  - i. **context information includes context entries**

As explained in [7.1], Maes discloses a context stack comprising “*context information... includ[ing] a plurality of context entries.*” Ex.1003, ¶110.

**ii. identifying ... context entries corresponding to the words**

As explained in [7.3], Maes discloses producing “decoded text” (or, script) (“*one or more words*”) processed from the spoken utterance (“*the natural language utterance*”). Further, as explained below, Maes teaches or suggests the “*identifying ...*” step. Ex.1003, ¶111.

Referencing the in-vehicle example, Maes describes its system identifying words in a recognized utterance from contexts—such as by matching the utterance “turn it on” to the context “radio”:

The dialog manager would ... receive the results of the recognized events associated with the spoken utterance “turn it on” and the gesture of pointing to the radio. **Based on these events, the dialog manager does a search of the existing applications, transactions or “dialogs,” or portions thereof [stored on the context stack], with which such an utterance [including the accompanying lip movement] and gesture could be associated.**

Ex.1005, 7:65–8:4.

**[T]his recognized spoken utterance event is stored on the context stack.** Then, when the recognized gesture event (e.g., pointing to the radio) is received, the dialog

manager takes this event and **the previous spoken utterance event stored on the context stack and makes a determination that the user intended to have the radio turned on.**

Ex.1005, 8:37–42; Ex.1003, ¶113.

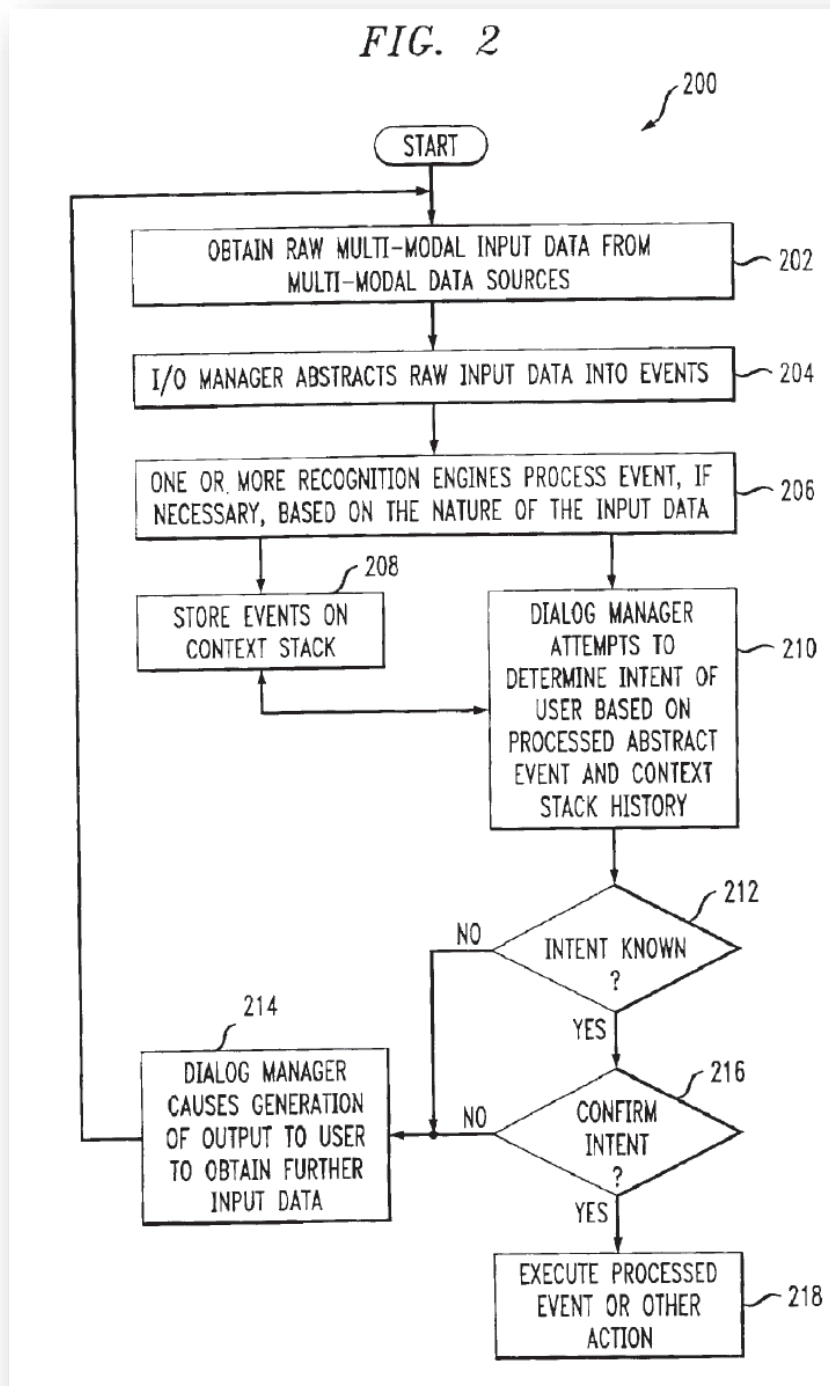
In this example, the results of the recognized events comprise decoded text (or, script) (“*the one or more words*”) associated with the spoken utterance “turn it on” with the accompanying lip movement. Further, in the above example, “this event” corresponds to the current event of the user pointing to the radio, which occurred few seconds after “the previous spoken utterance event” of the user saying “turn it on.” Data for both events (“*the plurality of context entries*”) are stored on the context stack because Maes discloses “**all** the variable, states, input, output and queries to the backend ... performed in the context of the dialog and any extraneous event that occurs during the dialog” are stored on the context stack. Ex.1005, 37:55–62. Maes clarifies that “the dialog” includes “**conversational dialog comprising speech** and other multi-modal I/O.” Ex.1005, 36:49–54; Ex.1003, ¶114.

Therefore, Maes discloses the dialog manager associating the results of the recognized events comprising decoded text (or, script) (“*the one or more words*”) with organized/sorted contexts in a context stack such as context stack 817 in Figure 8 or context stack 20 in Figure 1 corresponding to an active dialog which serves as

the “*identifying ...*” step. Accordingly, the associated context stored in the context stack corresponding to the user’s intent in the active dialog constitutes the “*one or more context entries that correspond to the one or more words.*” See also Ex.1005, 37:53–55. Ex.1003, ¶115.

Therefore, Maes teaches or suggests the “*identifying ...*” step for at least this reason. Ex.1003, ¶116.

Further, Maes Figure 2, which depicts “a flow diagram illustrating a referential ambiguity resolution methodology performed by a multi-modal conversational computing system” used in determining the user’s intent additionally teaches or suggests the “*identifying ...*” step. Ex.1005, 3:24–28.



Ex.1005, Figure 2. Maes explains that “[i]n step 208, the recognized events ... are stored in a storage unit referred to as the context stack 20. The **context stack** is used

to create a history of interaction between the user and the system ... **to assist the dialog manager 18 in making referential ambiguity resolution determinations** when determining the user's intent. Next, in step 210, the system 10 attempts to **determine the user intent based on the current event and the historical interaction information stored in the context stack ...**.” Ex.1005, 7:33–45.

Therefore, Maes teaches or suggests the “*identifying ...*” step for this additional reason. Ex.1003, ¶118.

For the above reasons, the associated context(s) stored in the context stack corresponding to the user's intent in the active dialog constitutes “*the plurality of context entries*” that “match[]” (“*correspond to*”) the decoded text (or, script) (“*the one or more words*”) associated with the spoken utterance, e.g., “turn it on” with the accompanying lip movement. Ex.1003, ¶119.

Dr. Houh explains Maes teaches identifying matches between associated context(s) stored in the context stack corresponding to the user's intent in the active dialog (or, simply contexts stored on the context stack) and results of recognized events (or, simply transcribed user utterances). Ex.1003, ¶120.

To the extent PO contends that Maes does not teach or suggest the “*identifying ...*,” step, Coffman does. See § VII.B.; Ex.1003, ¶121.

Coffman explains that context stack entries are searched to identify a context matching transcribed user input. “**Completion/search for the active context is**

**performed from context to context down the stack. That is, new queries or arguments are compared by the dialog engine by going down the stack until an acceptable match is obtained ... .”** Ex.1012, 42:1–3. Coffman further explains that searching for the context includes comparing transcribed user inputs against information stored in the context stack:

Based on the modality (pointer, keyboard, file, speech), the task dispatcher 402 redirects the stream to the appropriate conversational subsystems or conventional subsystem with **speech inputs being transcribed/understood. The output of these subsystems is run down the context stack 405 to extract the active query and complete it.**

Ex.1012, 41:2–5. Thus, in Coffman, the outputs of conversational subsystems or conventional subsystems which constitute transcribed user input, are compared against data stored in the context stack until a match is found.

Completion/search for the active context is performed from context to context down the stack. That is, **new queries or arguments are compared by the dialog engine by going down the stack until an acceptable match is obtained ....**

Ex.1012, 42:1–3. Thus, Coffman’s teachings of searching the context stack until an acceptable match is found between the transcribed user inputs and data stored on the

context stack serve as the “*identifying ...*” functionality. Indeed, Coffman teaches updating the context stack in a manner such that “[w]hen a message is returned to the user, the context is updated and then pushed to the top of the context stack ... under the active context. The active context is updated to inform of the existence of a returned value.” Ex.1012, 42:7–9. This additionally confirms that Coffman provides the “*identifying*” functionality. Ex.1003, ¶122.

Accordingly, the decoded script produced as a result of processing the user’s spoken utterance, as Maes teaches, would be compared against data stored on the context stack, per Coffman’s teachings, to “*identify[] one or more context entries that correspond to the one or more words.*” Ex.1012, 41:1–5, 40:15–19; Ex.1005, 7:60–8:4, 8:37–42, and 37:53–55. See § VII.B. Ex.1003, ¶122.

**f. [7.4.1] wherein identifying the one or more context entries comprises: comparing the plurality of context entries to the one or more words;**

As part of the “*identifying ...*” step explained in [7.4.0], Maes discloses that upon searching the context stack, the associated context stored in the context stack corresponding to the active dialog is provided by the dialog manager when such context “could be associated” with the “[*natural language*] utterance,” which therefore constitutes “*comparing*” data stored on the context stack (“*the plurality of context entries*”) “to” “decoded text” (or, script) (“*the one or more words*”) processed from the spoken utterance. Ex.1003, ¶125.

For example, Maes discloses:

Thus, the dialog manager must first determine the user's intent based on the current event and, if available, the historical information (e.g., past events) stored in the context stack. For instance, returning to the vehicle example, the user may say "turn it on," while pointing at the vehicle radio. The dialog manager would therefore receive the results of the recognized events associated with the spoken utterance "turn it on" and the gesture of pointing to the radio. **Based on these events, the dialog manager does a search of the existing applications, transactions or "dialogs," or portions thereof [stored on the context stack], with which such an utterance and gesture could be associated.**

Ex.1005, 7:60–8:4, *see also* 8:37–42. Therefore, Maes discloses that the "*identifying ... comprises: comparing the plurality of context entries to the one or more words.*"

Ex.1003, ¶126.

To the extent that PO contends that Maes does not disclose the above step, Coffman (incorporated by Maes) does:

Based on the modality (pointer, keyboard, file, speech), the task dispatcher 402 redirects the stream to the appropriate conversational subsystems or conventional subsystem with **speech inputs being transcribed/understood. The output of these**

**subsystems is run down the context stack 405 to extract the active query and complete it.**

Ex.1012, 41:2–5. Thus, in Coffman, the outputs of conversational subsystems or conventional subsystems which constitute one or more words transcribed/determined from the user’s spoken utterance, are compared against data stored in the context stack until a match is found. Thus, Coffman’s teachings of searching the context stack until an acceptable match is found between one or more words transcribed/determined from the user’s spoken utterance and data stored on the context stack serve as the “*identifying ... comprises: comparing the plurality of context entries to the one or more words.*” Ex.1003, ¶127.

Coffman explains that identifying context entries stored on the context stack includes “*comparing*” data stored on the context stack to one or more words transcribed/determined from the user’s spoken utterance. “**Completion/search for the active context is performed from context to context down the stack.** That is, new queries or arguments [“the one or more words”] processed from the utterance are compared [“comparing”] by the dialog engine by going down the stack [“the plurality of context entries”] until an acceptable match is obtained ... .” Ex.1012, 42:1–3; see also § VII.B. Ex.1003, ¶128.

**g. [7.4.2] [wherein identifying the one or more context entries comprises:] generating, based on the**

**comparison, one or more rank scores for individual context entries of the plurality of context entries; and**

As part of the “*identifying ...*” step in [7.4.0] and based on the outcome of the “*comparing ...*” step in [7.4.1], Maes teaches “*generating ... scores for individual context entries*” or otherwise scoring one or more candidate contexts (such as phonemes) stored on the context stack. Maes describes “phonemes” as “sub-phonetic or acoustic units of speech” (Ex.1005, 12:29–33) or equivalently as “portions [of dialogs]” (Ex.1005, 7:60–8:4). Ex.1003, ¶131.

Specifically, Maes discloses generating, for each considered phoneme in a given context, likelihood scores (indicating the likelihood that it was that particular phoneme that was spoken) based on the received audio information (and/or video information). Ex.1003, ¶132.

Referencing Figure 4, Maes discloses “[T]he probability module 416 in the audio information path ... labels the acoustic feature vectors with one or more phonemes.” Ex.1005, 18:46–48. Maes explains this in greater detail:

**After the acoustic feature vectors ... are extracted, the probability module labels the extracted vectors with one or more previously stored phonemes which ... are sub-phonetic or acoustic units of speech. The module may also work with lefemes, which are portions of phones [sic: phonemes] in a given context. Each phoneme associated with one or more feature vectors**

**has a probability associated therewith indicating the likelihood that it was that particular acoustic unit that was spoken.**

Thus, the probability module yields likelihood scores for each considered phoneme in the form of the probability that, given a particular phoneme or acoustic unit (au), the acoustic unit represents the uttered speech characterized by one or more acoustic feature vectors A or, in other words,  $P(A|\text{acoustic unit})$ .”

Ex.1005, 12:29–43. Maes further describes that

[E]ach phoneme associated with one or more visual speech feature vectors has a probability associated therewith indicating the likelihood that it was that particular acoustic unit that was spoken in the video segment being considered. Thus, **the probability module yields likelihood scores for each considered phoneme in the form of the probability** that, given a particular phoneme or acoustic unit (au), the acoustic unit represents the uttered speech characterized by one or more visual speech feature vectors V or, in other words,  $P(V|\text{acoustic unit})$ . ...

*See generally* Ex.1005, 18:46–60; Ex.1003, ¶133.

Further, as Dr. Hoh explains, the outcome of the “*identifying ...*” step (*see* [7.4.0]) would generate one or more candidate contexts from the context stack

(Ex.1003, ¶120), and as such, the candidate contexts constitute “*individual context entries of the plurality of context entries*” in the context stack. Ex.1003, ¶134.

In Maes, “likelihood scores” (“*one or more rank scores*”) generated for “each ... phoneme” / candidate context (“*individual context entries of the plurality of context entries*”) is “*based on*” “associat[ing]” [i.e., “*the comparison*”] words of the user’s spoken utterance against information stored in the context stack. Ex.1005, 7:60–8:4, 8:37–42, 12:35–37; *see also* [7.4.1] explaining “*the comparison*” / the “*comparing ...*” step). Accordingly, Maes teaches calculation of probabilities/likelihood scores (e.g., “*generating ... one or more rank scores*”) for each considered phoneme in a given context (e.g., “*individual context entries of the plurality of context entries*”). Ex.1003, ¶135.

**h. [7.4.3] [wherein identifying the one or more context entries comprises:] identifying, based on the one or more rank scores, the one or more context entries from among the plurality of context entries; and**

As explained in element [7.4.0] and [7.4.2], Maes teaches or suggests the “*identifying ... one or more context entries from among the plurality of context entries*” step and calculating “likelihood scores” (“*the one or more rank scores*”) for each considered phoneme. Ex.1003, ¶138.

As Dr. Houh explains, the “*identifying ...*” step is “*based on*” “likelihood scores” (“*one or more rank scores*”) because Maes teaches “likelihood scores” (“*one*

*or more rank scores*”) of a considered/identified phoneme “indicating the likelihood that **it was that particular acoustic unit that was spoken.**” Ex.1005, 12:35–37. Ex.1003, ¶139 (citing Ex.1005, 12:35–47).

Indeed, Maes discloses using a “degree of confidence” (e.g., “*one or more rank scores*”) to determine that the “the [selected] application” (e.g., *one or more context entries from among the plurality of context entries*) “is the one which will effectuate the users desire” and accordingly, “the next step of the multi-modal dialog ... of that application based on the multi-modal input” is executed. Ex.1005, 8:19–21 and 8:8–11; *see also* Ex.1005, 7:60–8:4 (disclosing applications as an example of an entry stored on the context stack). Accordingly, Maes teaches selecting a particular application (e.g., “*identifying the one or more context entries*”) using (“*based on*”) a “degree of confidence” (e.g., “*one or more rank scores*”); Ex.1003, ¶140.

Therefore, Maes teaches or suggests this element. Ex.1003, ¶.

- i. **[7.5] determining based on the determined one or more words and the context information, the command or the request associated with the natural language utterance.**

As explained in connection with elements [7.3] and [7.1], in Maes, “decoded text” (or, script) constitutes “*the determined one or more words*” processed from the spoken utterance and “historical information,” e.g., past input/output (I/O) events

generated previously in the context of a dialog and stored in the context stack constitutes “*the context information.*” And additionally, for the reasons below, Maes discloses this claim element. Ex.1003, ¶143.

**i. Determining the command/request associated with the natural language utterance**

Maes teaches or suggests “*determining ... the command or the request associated with the natural language utterance*” because Maes’s system provides functionality to “determine[] what information the user has, which inputs the user presents, and which application(s) should handle the user inputs [e.g., “the natural language utterance”].” Ex.1005, 36:49–59; Ex.1003, ¶144.

Specifically:

When an abstract event occurs, the DMA determines the target of the event ... [and] then launches the action associated to the **user’s query** [e.g., received as multi-modal data in the form of audio input data, video input data, as well as other types of input data]. The DMA function **handles multi-modal I/O events to: (1) determine the target application or dialog** (or portion of it); and (2) use past history and context to: (a) understand the intent of the user; (b) follow up with a dialog to disambiguate, complete, correct or confirm the understanding; (c) or, dispatch a task resulting from fall understanding of the intent of the user.

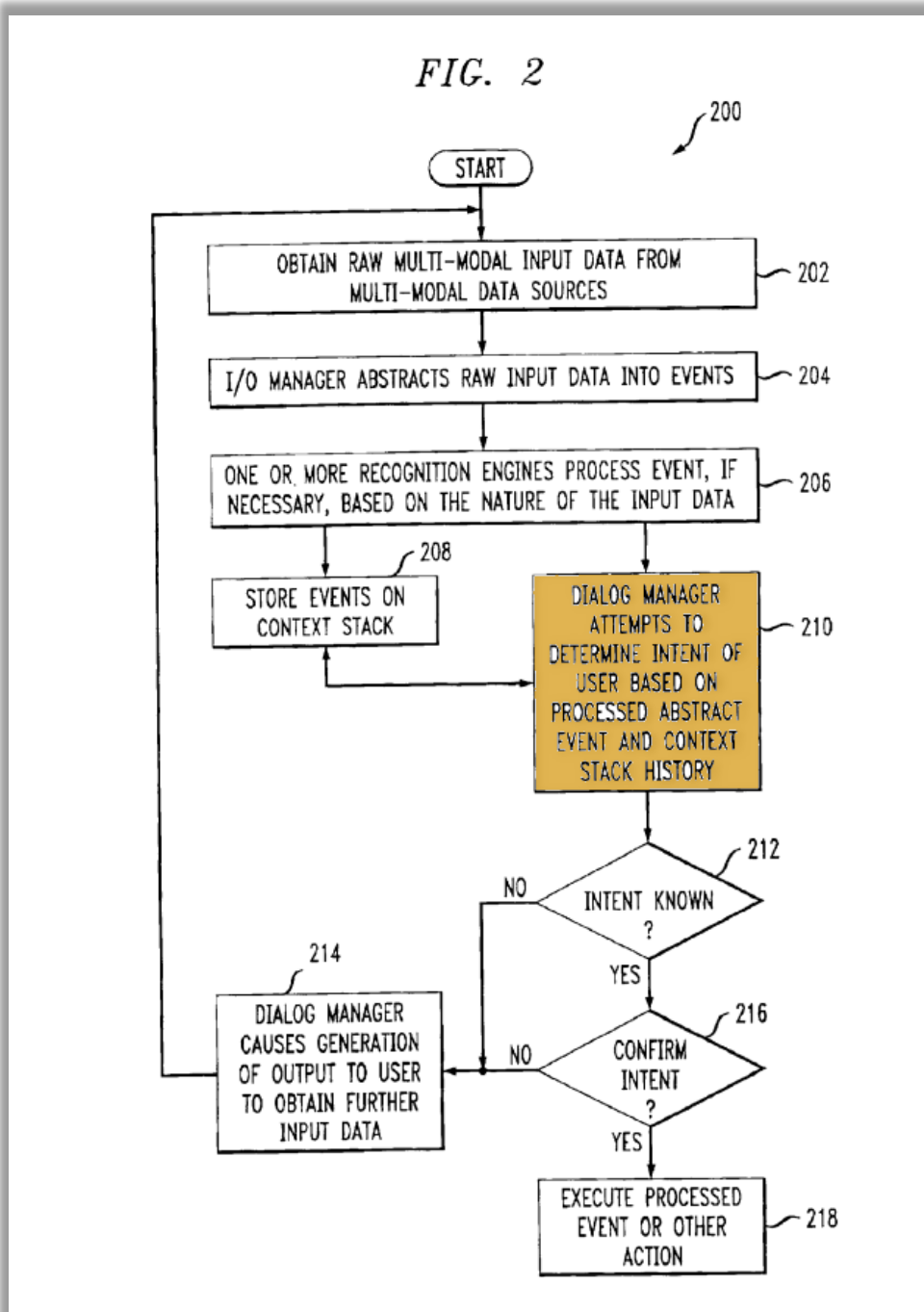
Ex.1005, 36:59–37:3; *see also* 4:6–22. Maes’s “**system ... recognizes voice-based commands** and ... gesture-based commands.” Ex. 1005, 2:1–4; *see also* Ex.1005,

4:58–59 (disclosing “user input commands received and processed by the system”).  
Ex.1003, ¶145.

**ii. Determining based on one or more words and the context information,**

In Maes, “*determining ...*” “a user’s query” (“*the command or the request*”) is “*based on*” “decoded text” (or, script) (“*the determined one or more words*”) which is produced using context stack data (“*context information*”), explained below. Ex.1003, ¶146.

Referencing the Figure 2 process, Maes describes “in step 210, **the system ... attempts to determine the user intent based on the current event and the historical interaction information stored in the context stack.**” Ex.1005, 7:40–42. In step 210, the “processed abstract event” is the “decoded text” (or, script) (“*the determined one or more words*”) and the “context stack history” constitutes “*the context information.*”



Ex.1005, Figure 2 (annotated); Ex.1003, ¶147.

Referencing the in-vehicle example, Maes states:

[T]he dialog manager must first determine the user's intent based on the current event and ... the historical

information (e.g., past events) stored in the context stack. For instance, returning to the vehicle example, the user may say “turn it on,” while pointing at the vehicle radio. The dialog manager would ... receive the results of the recognized events associated with the spoken utterance “turn it on” and the gesture of pointing to the radio. Based on these events, the dialog manager does a search of the existing applications, transactions or “dialogs,” or portions thereof [stored on the context stack], with which such an utterance and gesture could be associated.

Ex.1005, 7:65–8:4, *see also* 8:37–42. In this example, determining the user’s intent (*determining ... the command or the request associated with the natural language utterance*) is based on the results of the recognized events which comprise decoded text (or, script) (*“the one or more words”*) associated with the spoken utterance “turn it on” with the accompanying lip movement and the historical information (e.g., past events) stored in the context stack (*“the context information”*). Ex.1003, ¶148.

Maes further discloses that its system:

provides the capability to: (i) determine an object, application or appliance addressed by the user; (ii) determine the focus of the user and therefore determine if the user is actively focused on an appropriate application and, on that basis, to determine if an action should be taken; (iii) **understand queries** [*“determining ... the command or request”*] **based on ... said ... what** [*“the determined one or more words”*], what was the focus of the user when he gave a multi-modal query/command **and**

**what is the history of these commands** [*“the context information”*] and focuses ....

Ex.1005, 2:54–67; Ex.1003, ¶149.

Therefore, Maes teaches or suggests this claim element. Ex.1003, ¶150.

## 2. Claim 8

- a. **[8.0] The method of claim 7, further comprising: providing a prompt regarding the command or request, wherein the prompt is a prompt for additional information or a prompt for confirmation; and**
  - i. **Providing a prompt regarding the command or request**

Referencing the in-vehicle example and Figure 2, Maes explains:

Consider the case where the user says “turn it on,” but makes no gesture and provides no other utterance. In this case, assume that the dialog manager does not have enough input to determine the user intent (step 212 in FIG. 2) and thus implement the command. The dialog manager, in step 214, .. causes the **generation of an output to the user requesting further input data so that the user’s intent can be disambiguated.** ... [For example,] [t]he dialog manager ... **generates a predetermined question such as “what device do you want to have turned on?”** ....

*See generally*, Ex.1005, 8:43–65; Ex.1003, ¶152.

In the above example, providing the question “what device do you want to have turned on?” constitutes “*providing a prompt*” “*regarding*” “the user’s query” (“*the command or request*”), *e.g.*, relating to the spoken utterance of “turn it on.” *Id.*, Ex.1003, ¶153.

**ii. The prompt is a prompt for additional information or confirmation**

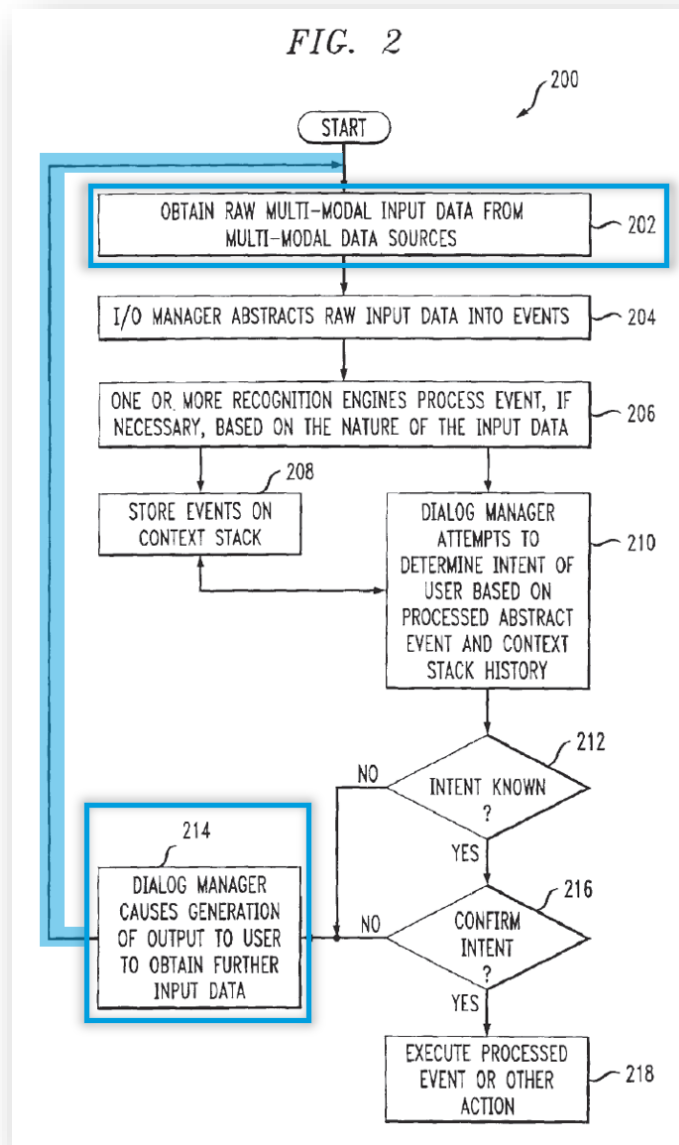
Additionally, continuing with the above example (Ex.1005, 8:43–65), the question (“*the prompt*”) “*is*” “output to the user requesting further input data” (“*a prompt for additional information ...*”). Ex.1003, ¶154.

Maes further explains that when an abstract input/output (I/O) event occurs, Maes’s system seeks “**confirmation**, disambiguation, correction, **more details**” (“*a prompt for additional information or ... confirmation*”) (“*a prompt for additional information or ... confirmation*”). Ex.1005, 36:59–63. Specifically, Maes’s system seeks a user’s follow-up input / feedback “to disambiguate, complete, correct or **confirm** the understanding.” Ex.1005, 36:64–37:2; Ex.1003, ¶155.

Maes discloses requesting additional information/confirmation with more disclosures:

The dialog manager 18 may also seek **confirmation** in step 216 from the user in the same manner as the request for **more information** (step 214) before executing the processed event ... . For example, the system may output “**do you want the radio turned on?**” To which the user may respond “yes.” ....

Ex.1005, 8:66–9:6. In this example, Maes’s system asking the user “do you want the radio turned on?” constitutes “*a prompt for additional information ...*”.



Ex. 1005, Figure 2 (annotated) (depicting “generation of output to user to obtain further input data” received in block 214); Ex.1003, ¶156.

- b. **[8.1] receiving, in response to providing the prompt, a non-speech input, wherein the command or request is determined further based on the non-speech input.**

Continuing with the in-vehicle system example discussed in [8.0] (*see* Ex.1005, 8:43–65), Maes teaches that upon hearing the question “what device do you want to have turned on?,” the user “could then point to the radio or say “the radio” thereby providing the dialog manager with the additional input data to disambiguate his request.” 8:58–61. Thus, Maes’s system receives (“*receiv[es]*”) in response to providing the question (“*the prompt*”), the user’s gesture of pointing to the radio or the lip movement accompanying the spoken utterance “the radio,” each of which serve as examples of “*a non-speech input.*” Furthermore, Maes teaches that the user’s query (“*the command or request*”), *e.g.*, relating to the spoken utterance of “turn it on” is disambiguated (“*determined further*”) “*based on*” the user’s gesture or the lip movement accompanying the spoken utterance (“*non-speech input*”). Ex.1003, ¶158.

**B. Ground 2: Maes and Coffman render obvious claims 7 and 8.**

To the extent one asserts that Coffman is not particularly incorporated by reference into Maes as if they were effectively the same document, claims 7 and 8 would have nonetheless been obvious in view of Maes and Coffman for the reasons articulated above in Ground 1. Moreover, as demonstrated above, a POSITA would have been motivated to combine the references and had a reasonable expectation of success without undue experimentation for such a combination as articulated in § VII.B. Ex.1003, ¶160.

**C. Ground 3: Kennewick and Weissman, in further view of Maes renders obvious claims 7 and 8.**

**1. Claim 7**

**a. Dialect is Collaterally Estopped from Relitigating the Obviousness of Claim 7**

In an *ex parte* reexamination involving the 957 Patent, the Patent Office notified PO that it would be using a reexamination certificate cancelling, among other claims, claim 7. *See* Reexam Ctrl No. 90/015,303 (Notice of Intent to Issue Ex Parte Reexamination Certificate – March 6, 2025). Claim 7 was cancelled as obvious under at least 35 U.S.C. § 103 based on, for example, U.S. Patent Publication No. 2004/0044516 to Kennewick et al. (“Kennewick”) (Ex.1032) in view of U.S. Patent No. 6,453,315 to Weissman (“Weissman”) (Ex.1025), each of which are prior art to the 957 Patent under at least 35 U.S.C. § 102(b). *See* Reexam Ctrl No. 90/015,303 (Final Rejection – June 27, 2024). PO is collaterally estopped from relitigating the unpatentability of claims 7-8 because the previously rejected claims were either identical (claim 7) or not materially different (claim 8) from the challenged claims in this IPR. *Google LLC v. Hammond Development Int’l*, 54 F.4th 1377, 1381-1382 (Fed. Cir. 2022).

Collateral estoppel applies to this proceeding, *Google*, 54 F.4th at 1381, and a final PTAB decision has preclusive effects. *See SynQor, Inc. v. Vicor Corp.*, 988 F.3d

1344 (Fed. Cir. 2021) (applying collateral estoppel based on the PTAB’s decision in a previous reexamination proceeding).

For collateral estoppel to apply, Petitioner need only demonstrate that “(1) the issue is identical to one decided in the first action; (2) the issue was actually litigated in the first action; (3) resolution of the issue was essential to a final judgment in the first action; and (4) PO had a full and fair opportunity to litigate the issue in the first action.” *Google*, 54 F.4th at 1381. All four elements are met here because (1) and (2) claim 7 was specifically at-issue in the *ex parte* reexamination proceeding, (3) the patentability of claim 7 was at-issue and Kennewick and Weissman were considered and used as a basis of rejection by the Patent Office and the claim was found to be invalid and (4) a final judgement was issued after PO failed to file an appeal brief.

Accordingly, Kennewick and Weissman render obvious claim 7.

## 2. Claim 8

- a. **[8.0] The method of claim 7, further comprising: providing a prompt regarding the command or request, wherein the prompt is a prompt for additional information or a prompt for confirmation; and**
- b. **[8.1] receiving, in response to providing the prompt, a non-speech input, wherein the command or request is determined further based on the non-speech input.**

Maes teaches or suggests all the elements of claim 8 for the reasons demonstrated above in Grounds 1 and 2. *See* [8.0] and [8.1] *above*; Ex.1003, ¶162.

It would have been obvious to a POSITA to combine Kennewick and Weissman with Maes's teachings regarding prompting for additional input or confirmation, including where such additional input or confirmation is received by means other than speech. The combination would have predictably resulted in Kennewick and Weissman's speech recognition processing functionalities enhanced to include follow-up prompting, including from non-speech data, per Maes, as claimed. Ex.1003, ¶163.

Kennewick, Weissman, and Maes are analogous art as the 957 Patent for two reasons. *First*, each are directed toward the same field of endeavor as the 957 Patent—computer-implemented and natural language processing systems that include a focus on the meaning of language (e.g., semantics), and are reasonably pertinent to a problem described by the 957 Patent: an environment for **reliably processing** a user's language queries, which is a problem of conventional systems identified in the 957 Patent. Ex.1001, Abstract, Ex.1001, 41:58-64; Ex.1032, [0003], [0009]; Ex.1025, Abstract, 2:17-49; Ex.1005, 41 :58–64; Ex.1003, ¶164.

Dr. Houh explains that a POSITA would have been motivated to combine Kennewick, Weissman and Maes in the manner described. As discussed above, a POSITA would have already understood and found it obvious based on Kennewick's disclosures and background knowledge that Kennewick's system sought means to reliably process user requests, and Maes discloses a technique that seeks a prompt

for additional information and/or confirmation: **“confirmation, disambiguation, correction, more details, etc., until the intent is unambiguous and fully determined”** Ex.1005, 36:59–63. Specifically, Maes discloses a technique that seeks a user’s follow-up input / feedback “to disambiguate, complete, correct or confirm the understanding.” Ex.1005, 36:64–37:2; Ex.1003, ¶165.

Kennewick confirms this understanding by demonstrating a similar technique—it explains that:

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

Ex.1032, [0131]; *see also* Ex.1032, [0145]. Because of the similarity of Maes’s techniques, a POSITA, seeking to further understand or implement Kennewick’s disclosures, would have appreciated that Maes adds certain additional functionality,

including multi-modal input, so a POSITA would have turned to Maes for its enhanced functionality and more detailed teachings. Ex.1003, ¶166.

A person of ordinary skill in the art would have had a reasonable expectation of success, as combining and Kennewick would have involved the straightforward application of conventional techniques and technologies without undue experimentation. Ex.1003, ¶167.

Accordingly, Kennewick and Weissman, in further view of Maes renders obvious claim 8. Ex.1003, ¶168.

## **IX. THE BOARD SHOULD NOT EXERCISE ITS DISCRETION TO DENY INSTITUTION**

### **A. The Board Should Not Deny Institution Under 35 U.S.C. § 325(d)**

Maes, relied in this Petition as a single-reference obviousness ground, was not used in rejecting the claims by the Examiner during prosecution of the 957 Patent. The Examiner was presented with Maes as **one of 100s of references** submitted as part of a 28-page IDS submission by Patent Owner although the Examiner failed to perform a substantive evaluation of Maes. See Ex.1001, 1-6. Coffman, incorporated by reference in Maes (Ex.1005, 9:30–37) is also relied upon in this Petition to the extent Patent Owner contends that Maes does not teach or suggest certain claim limitations. A “relative” of Coffman—the US National stage of Coffman that issued as US Pat. No. 7,137,126 (“Coffman126” or Ex.1013)—was listed on an IDS, but

was also not evaluated substantively and was not the basis of a rejection. *See CrowdStrike, Inc. v. Webroot Inc.*, IPR2023- 00126, Paper 9 at 14 (PTAB May 5, 2023) (“[the Board] cannot determine the extent to which [cited reference] was evaluated during the examination because it was never used in a rejection” and because “[cited reference] was not the basis of a rejection, there is no overlap of arguments.”); *Scientific Design Co. Inc. v. Shell Oil Co.*, IPR2022-00158, Paper 7 at 24-26 (PTAB Apr. 4, 2022) (declining to exercise discretion under § 325(d) based on “credit[ing] Petitioner’s argument that although the references ‘were cited in an IDS, together with 111 references, none of them was used in any rejections, or otherwise addressed by the examiner’”); *Bowtech, Inc. v. MCP IP, LLC*, IPR2019-00379, Paper 14 at 18 (PTAB Jul. 3, 2019) (“[the Board has] consistently held that a reference that was neither applied against the claims nor discussed by the Examiner does not weigh in favor of exercising our discretion under § 325(d).”) The first part of the *Advanced Bionics* analysis is therefore not satisfied.

Nonetheless, Petitioner respectfully submits that the Examiner materially erred by not substantively evaluating the Maes reference (either alone or in conjunction with Coffman) against the 957 Patent and further erred by failing to reject any claims as obvious over Maes as highlighted by the strength of Petitioner’s Proposed Ground. Where, as here, a reference was never “substantively discussed by the Examiner[,]” “a petitioner’s showing that the challenged claims are

unpatentable over the asserted prior art may itself be evidence of material error by the Office during prosecution.” *Quasar Science LLC v. Colt International Clothing, Inc. d/b/a Colt LED*, IPR2023- 00611, Paper 10 at 14 (PTAB Oct. 10, 2023).

Further, the Examiner did not have the opportunity to review Dr. Houh’s detailed expert testimony supporting this petition. The Board has found the *Becton* factors favor institution where even a single noncumulative secondary reference distinguished the prior-art ground, regardless of other overlapping references. *Oticon Medical AB et al. v. Cochlear Limited*, IPR2019-00975, Paper 15 at 9-20 (P.T.A.B. Oct. 16, 2019) (precedential). The Board has declined to deny institution where Petitioners relied upon prior art cited on the face of the challenged patent but “not applied by the examiner ... in any rejection of claims.” *Comcast Cable Communs., LLC v. Promptu Sys. Corp.*, Case No. IPR2018-00342, Paper 13 at 17 (P.T.A.B. July 19, 2018).

**B. The Board Should Not Deny Institution Under 35 U.S.C. § 314(a)**

**1. Factor 1: Stay**

Factor is **neutral or weights against discretionary denial** because no request for stay has been filed in the litigation involving BOA, and Microsoft is not a defendant in that litigation. *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 15, 12 (May 13, 2020) (informative) (“*Fintiv*”); *Google, LLC v. Parus Holdings Inc.*, IPR2020-00847, Paper 9 at 12 (PTAB Oct. 21, 2020).

## 2. Factor 2: Trial Date

Microsoft is not a defendant in litigation involving this patent so this factor **weighs heavily against denial**. However, BOA—an RPI—is. Trial for BOA will not occur before January 26, 2026. Any FWD can be expected before the end of September 2026, which is eight months after the currently scheduled trial. “[A] court’s general ability to set a fastpaced schedule is not particularly relevant ... where, like here, the forum itself has not historically resolved cases so quickly.” *In re Apple Inc.*, 979 F.3d 1332, 1344 (Fed. Cir. 2020). Indeed, there are **13 other trials** before Judge Gilstrap that are currently scheduled for January 26, 2026, so it appears likely that the trial date will slip. Ex.1029. Moreover, the trial date is 10 months away and much can change during this time. *Dish Network v. Broadband iTV*, IPR2020-01280, Paper 17 at 16 (PTAB Feb. 4, 2021) (“We cannot ignore the fact that the currently scheduled trial date is more than nine months away and much can change during this time.”); *see also NetNut Ltd. v. Bright Data Ltd.*, IPR2021-01492, Paper 12 (PTAB Mar. 21, 2022) (granting institution when trial was to be held six months before the FWD).

Accordingly, this factor is **neutral or weighs against** discretionary denial.

## 3. Factor 3: Parallel Proceeding

Factor 3 is **neutral or weighs against** discretionary denial. Microsoft is not a defendant for this patent. And in the related litigation involving BOA, claim

construction briefing does not start until June 17, 2025, and the Claim Construction hearing is not scheduled until July 29, 2025. The investment by the parties to date has been relatively minimal.

#### **4. Factor 4: Overlapping Issues**

Consistent with *Sotera Wireless* Petitioner and BOA stipulate that if the PTAB institutes *inter partes* review of this proceeding, neither Petitioner nor BOA will pursue in any related district court case the grounds that were raised in this Petition, nor any other grounds that could have been reasonably raised in this Petition. *Sotera Wireless, Inc. v. Masimo Corp.*, IPR2020-01019, Paper 12 at 13-14 (PTAB Dec. 1, 2020) (precedential as to § II.A).

Factor 4 **weighs strongly against** discretionary denial.

#### **5. Factor 5: Overlapping Parties**

Factor 5 is **neutral or weighs against** discretionary denial. Microsoft provides software accused of infringement of the challenged patent in related litigation involving BOA. As *Fintiv* demonstrates, this factor should play a role only where a Petitioner is unrelated to a defendant in a district court proceeding. See *Fintiv*, IPR2020-00019, Paper 11 at 13–14 (PTAB Mar. 20, 2020).

#### **6. Factor 6: Strength/Other Considerations**

“[I]n circumstances where the Board determines that the other *Fintiv* factors 1–5 do not favor discretionary denial, the Board shall decline to discretionarily deny

under *Fintiv* without reaching the compelling merits analysis.” *CommScope Techs. LLC v. Dali Wireless, Inc.*, IPR2022-01242, Paper No. 23 at 4–5 (PTAB Feb. 27, 2023). For at least this reason, the Board need not reach the question on Factor 6.

Even so, the merits of the Petition are particularly compelling—for example, Maes (and, by incorporation Coffman) is a strikingly similar disclosure to that of the 957 Patent, is directed at the same problem, proposes the same solutions as 957 Patent, and discloses nearly identical functionality as the claimed “*method of processing a natural language utterance*.” The evidence presented, if unrebutted, would lead to a conclusion that one or more claims are unpatentable by a preponderance of the evidence.

## **X. CONCLUSION**

Claims 7 and 8 of the 957 Patent are unpatentable for the reasons discussed above.

Dated: March 21, 2025

Respectfully Submitted,

/Scott M. Border/

Scott M. Border

Reg. No. 77,744

Winston & Strawn LLP

1901 L Street, N.W.

Washington, D.C. 20036

T: 202-282-5054

[sborder@winston.com](mailto:sborder@winston.com)

**CERTIFICATE OF COMPLIANCE**

I hereby certify that this brief complies with the type-volume limitations of 37 C.F.R. § 42.24, because it contains 11,749 words (as determined by the Microsoft Word word-processing system used to prepare the brief and including annotated figures), excluding the parts of the brief exempted by 37 C.F.R. § 42.24.

Dated: March 21, 2025

Respectfully Submitted,

/Scott M. Border/

Scott M. Border

Reg. No. 77,744

Winston & Strawn LLP

1901 L Street, N.W.

Washington, D.C. 20036

T: 202-282-5054

[sborder@winston.com](mailto:sborder@winston.com)

**CERTIFICATE OF SERVICE**

Pursuant to 37 C.F.R. § 42.6(e), I hereby certify that on this 21st day of March, 2025, I caused to be served a true and correct copy of the foregoing and any accompanying exhibits by Federal Express Priority Overnight on the following:

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

A courtesy copy of this Petition and supporting material was also served on litigation counsel for Patent Owner via email:

Garland Stephens (garland@bluepeak.law)  
Justin Constant (justin@bluepeak.law)  
Richard Koehl (richard@bluepeak.law)  
Kate Falkenstien (kate@bluepeak.law)  
Heng Gong (heng@bluepeak.law)  
**BLUE PEAK LAW GROUP LLP**  
3139 West Holcombe Blvd. PMB 8160  
Houston, TX 77025

Dated: March 21, 2025

Respectfully Submitted,

/Scott M. Border/

Scott M. Border

Reg. No. 77,744

Winston & Strawn LLP

1901 L Street, N.W.

Washington, D.C. 20036

T: 202-282-5054

[sborder@winston.com](mailto:sborder@winston.com)

**CLAIM LISTING**

| <b>Claim 7</b> |                                                                                                                                                                                                           |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [7.0]          | A computer-implemented method of processing a natural language utterance, the method comprising:                                                                                                          |
| [7.1]          | generating a context stack comprising context information that corresponds to a plurality of prior utterances, wherein the context stack includes a plurality of context entries;                         |
| [7.2]          | receiving, by one or more computer processors, the natural language utterance, wherein the natural language utterance is associated with a command or is associated with a request;                       |
| [7.3]          | determining one or more words of the natural language utterance by performing speech recognition on the natural language utterance;                                                                       |
| [7.4.0]        | identifying, from among the plurality of context entries, one or more context entries that correspond to the one or more words, wherein the context information includes the one or more context entries, |
| [7.4.1]        | wherein identifying the one or more context entries comprises: comparing the plurality of context entries to the one or more words;                                                                       |
| [7.4.2]        | [wherein identifying the one or more context entries comprises:] generating, based on the comparison, one or more rank scores for individual context entries of the plurality of context entries; and     |
| [7.4.3]        | [wherein identifying the one or more context entries comprises:] identifying, based on the one or more rank scores, the one or more context entries from among the plurality of context entries; and      |
| [7.5]          | determining based on the determined one or more words and the context information, the command or the request associated with the natural language utterance.                                             |
| <b>Claim 8</b> |                                                                                                                                                                                                           |
| [8.0]          | The method of claim 8, further comprising: providing a prompt regarding the command or request, wherein the prompt is a prompt for additional information or a prompt for confirmation; and               |
| [8.1]          | receiving, in response to providing the prompt, a non-speech input, wherein the command or request is determined further based on the non-speech input                                                    |