

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

SOUNDCLEAR TECHNOLOGIES, LLC,
Patent Owner.

Case No. IPR2025-01177
U.S. Patent No. 11,244,675

PETITION FOR *INTER PARTES* REVIEW

TABLE OF CONTENTS

I.	Introduction.....	1
II.	The '675 Patent (Ex. 1001).....	1
III.	Level of Ordinary Skill in the Art	5
IV.	Claim Construction.....	5
V.	Ground 1: Claims 1-7 Are Obvious over Raitio and Fahlman.....	5
	A. Raitio (Ex. 1006)	5
	B. Fahlman (Ex. 1007).....	11
	C. Independent Claims	13
	1. Claim 1	13
	a. [1preamble]: “An output-content control device comprising:”	14
	b. [1a]: “a voice acquiring unit configured to acquire a voice spoken by a user;”	16
	c. [1b]: “a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”	18
	d. [1c]: “an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;”	24
	e. [1d]: “a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and”	27

f.	[1e]: “an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information,”	30
g.	[1f]: “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word, and to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”	33
i.	Motivation to Combine Raitio and Fahlman.....	35
ii.	Expectation of Success.....	36
2.	Claim 6.....	37
a.	[6preamble]: “An output-content control method comprising:”	37
b.	[6a]: “acquiring a voice spoken by a user;”.....	39
c.	[6b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”	39
d.	[6c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”	40
e.	[6d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”	40

- f. [6e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and” 40
 - g. [6f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.” 40
- 3. Claim 7 40
 - a. [7preamble]: “A non-transitory storage medium that stores an output-content control program that causes a computer to execute:” 40
 - b. [7a]: “acquiring a voice spoken by a user;” 41
 - c. [7b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;” 42
 - d. [7c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;” 42
 - e. [7d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and” 42
 - f. [7e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and” 42
 - g. [7f]: “generating, when the voice is determined to be the second voice, a second output sentence

	which includes all of the intention information and the content information.”	42
D.	Dependent Claims	43
1.	Claim 2: “The output-content control device according to claim 1, wherein the notification information includes type information indicating a type of the content information, and the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be [re]placed with the other word based on a word included in the content information and the type information that is associated with the content information.”	43
2.	Claim 3: “The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.”	46
3.	Claim 4: “The output-content control device according to claim 3, wherein the relationship information further includes information of type information indicating a type of the first word, and the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word.”	47
4.	Claim 5: “The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the	

	predetermined voice when the voice spoken by the user is a whisper.”	49
VI.	Ground 2: Claims 1-7 Are Obvious over Ocampo and Fahlman.....	50
A.	Ocampo (Ex. 1005)	50
B.	Independent Claims	55
1.	Claim 1	55
a.	[1preamble]: “An output-content control device comprising:”	55
b.	[1a]: “a voice acquiring unit configured to acquire a voice spoken by a user;”	56
c.	[1b]: “a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”	58
d.	[1c]: “an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;”	61
e.	[1d]: “a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and”	65
f.	[1e]: “an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information,”	67
g.	[1f]: “wherein the output-content generating unit is further configured to generate, when the voice is	

	determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word, and to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”	72
i.	Motivation to Combine Ocampo and Fahlman	73
ii.	Expectation of Success	74
2.	Claim 6	75
a.	[6preamble]: “An output-content control method comprising:”	75
b.	[6a]: “acquiring a voice spoken by a user;”	76
c.	[6b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”	76
d.	[6c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”	77
e.	[6d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”	77
f.	[6e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and”	77

g.	[6f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”	77
3.	Claim 7	78
a.	[7preamble]: “A non-transitory storage medium that stores an output-content control program that causes a computer to execute:”	78
b.	[7a]: “acquiring a voice spoken by a user;”	78
c.	[7b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”	79
d.	[7c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”	79
e.	[7d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”	79
f.	[7e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and”	79
g.	[7f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”	80
C.	Dependent Claims	80

1.	Claim 2: “The output-content control device according to claim 1, wherein the notification information includes type information indicating a type of the content information, and the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be [re]placed with the other word based on a word included in the content information and the type information that is associated with the content information.”	80
2.	Claim 3: “The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.”	83
3.	Claim 4: “The output-content control device according to claim 3, wherein the relationship information further includes information of type information indicating a type of the first word, and the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word.”	84
4.	Claim 5: “The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.”	86
VII.	Mandatory Notices.....	87
A.	Real Parties-in-Interest Under 37 C.F.R. § 42.8(b)(1).....	87

B.	Related Matters Under 37 C.F.R. § 42.8(b)(2)	87
C.	Lead and Backup Counsel Under 37 C.F.R. § 42.8(b)(3).....	88
D.	Service Information Under 37 C.F.R. § 42.8(b)(4).....	89
VIII.	Standing	90
IX.	Conclusion	90

TABLE OF AUTHORITIES

Page(s)

Cases

<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.</i> , 868 F.3d 1013 (Fed. Cir. 2017)	5
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (en banc)	5

Statutes

35 U.S.C. § 102(a)(1).....	5, 11, 49
35 U.S.C. § 103	1

Rules

37 C.F.R. § 42.8(b)(1).....	87
37 C.F.R. § 42.8(b)(2).....	87
37 C.F.R. § 42.8(b)(3).....	88
37 C.F.R. § 42.8(b)(4).....	88

EXHIBITS

Exhibit	Description
1001	U.S. Patent No. 11,244,675 to Naganuma (“the ’675 patent”)
1002	Prosecution History of U.S. Patent No. 11,244,675
1003	Declaration of Stuart J. Lipoff (“Lipoff”)
1004	CV of Stuart J. Lipoff
1005	U.S. Patent Application Publication No. 2018/0122361 to Silveira Ocampo (“Ocampo”)
1006	U.S. Patent Application Publication No. 2017/0358301 to Raitio et al. (“Raitio”)
1007	U.S. Patent No. 5,960,080 to Fahlman et al. (“Fahlman”)
1008	RESERVED
1009	<i>SoundClear Technologies LLC v. Google LLC</i> , Case No. 1:24-cv-1281, Complaint for Patent Infringement (E.D. Va. July 25, 2024)
1010	<i>SoundClear Technologies LLC v. Google LLC</i> , Case No. 3:24-cv-00540, Memorandum Opinion (E.D. Va. Apr. 18, 2025)
1011	U.S. Patent Application No. 11/586,862
1012	U.S. Patent Application No. 11/638,251

I. Introduction

Petitioner Google LLC requests *inter partes* review of claims 1-7 of U.S. Patent No. 11,244,675 (“the ’675 patent”) (Ex. 1001), assigned to Patent Owner SoundClear Technologies LLC, and cancellation of the same based on the following grounds:

Grounds	Claims	Basis	Prior Art
1	1-7	§ 103	Raitio and Fahlman
2	1-7	§ 103	Ocampo and Fahlman

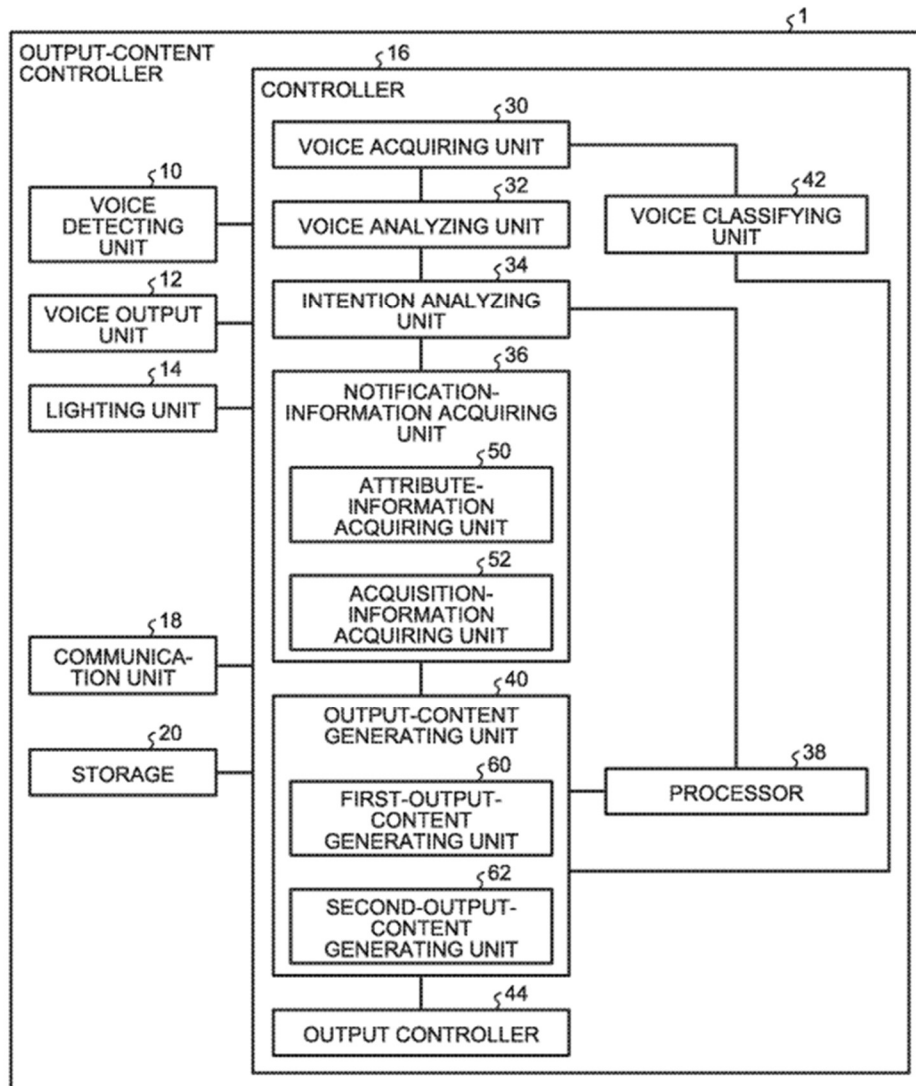
II. The ’675 Patent (Ex. 1001)

The ’675 patent, filed March 7, 2019, and issued February 8, 2022, claims priority to Japanese Patent Application No. JP2018-044598 filed March 12, 2018. Ex. 1001, cover (22), (30), (45).¹

The ’675 patent discloses generating responses to spoken user inputs. Ex. 1001, Abstract, 1:44-2:23. Figure 2 below illustrates an “output-content controller” (also called an “output-content control device”) for receiving and processing the spoken user input, and generating a response based on the input. Ex. 1001, Abstract, FIG. 2.

¹ Petitioner evaluates the claims under AIA standards.

FIG.2



Ex. 1001, FIG. 2.

The device includes a “voice detecting unit 10” configured to receive a user’s spoken “wish[]” and a “controller 16” comprised of various units that allow the device to “detect intention information indicating what kind of information” the user wishes to acquire; classify the voice as, e.g., a whisper or non-whisper; and

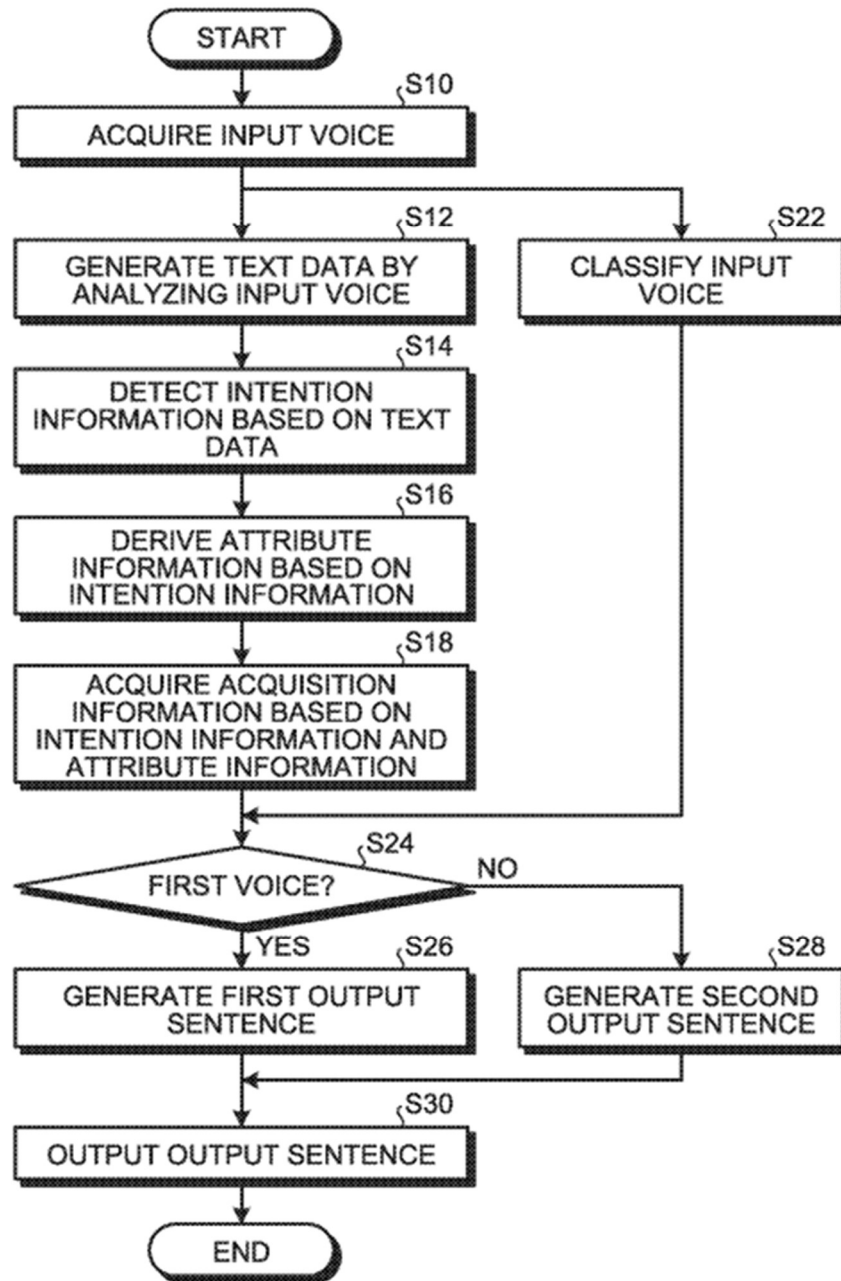
generate an output based on the voice's classification and the user's wish.

Ex. 1001, Abstract, 3:18-22, 3:27-31, 3:55-67, 4:15-24, 9:58-10:8, 11:5-21, FIG. 2.

For example, an "intention analyzing unit" may "extract[] ... intention information" from text data corresponding to the user's voice input using "a natural language processing" (Ex. 1001, 4:15-27); a "voice classifying unit" may classify a user's voice "V1" as a "first voice V1A when the voice V1 is determined to be a whisper" and "classif[y] the voice V1 [as] the second voice V1B when the voice V1 is determined not to be a whisper." (Ex. 1001, 9:65-10:4); a "notification-information acquiring unit 36 acquires notification information that is a content of information to be given to the user H based on the intention information" (Ex. 1001, 5:4-6); and an "output-content-generating unit" can generate an output sentence that responds to the user's wish using a first output sentence, or a second output sentence that replaces one or more words from the first output sentence, depending on whether the user's input is classified as voice V1A (whisper) or V1B (non-whisper) (Ex. 1001, 17:1-9). Figure 7 illustrates a process for generating a first or second output sentence based on voice classification and user intent.

Ex. 1001, 16:43-17:20, FIG. 7.

FIG.7



Ex. 1001, FIG. 7.

III. Level of Ordinary Skill in the Art

A person of ordinary skill in the art (“POSITA”) in the field of the ’675 patent would have had at least a bachelor’s degree in electrical engineering or related discipline, and at least two years of experience in voice or audio processing. Relevant work experience can substitute for formal education and additional education could substitute for work experience. Lipoff, ¶58.

IV. Claim Construction

The Board construes claims using the standards set forth in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc). Claims should be construed only to the extent necessary to resolve a controversy. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017). No terms need construction to resolve the unpatentability issues in this Petition.²

V. Ground 1: Claims 1-7 Are Obvious over Raitio and Fahlman

A. Raitio (Ex. 1006)

Raitio, filed September 15, 2016, and published December 14, 2017, is prior art to the ’675 patent under 35 U.S.C. § 102(a)(1). Raitio, cover (22), (45).

² Petitioner may raise indefiniteness and/or claim constructions in other forums.

Such challenges do not preclude the Board’s ability to resolve patentability.

Raitio discloses systems and processes for detecting and/or providing a whispered speech response. Figure 2A below “is a block diagram illustrating a portable multifunction device implementing the client-side portion of a digital assistant in accordance with some embodiments.” Raitio, [0009], FIG. 2A.

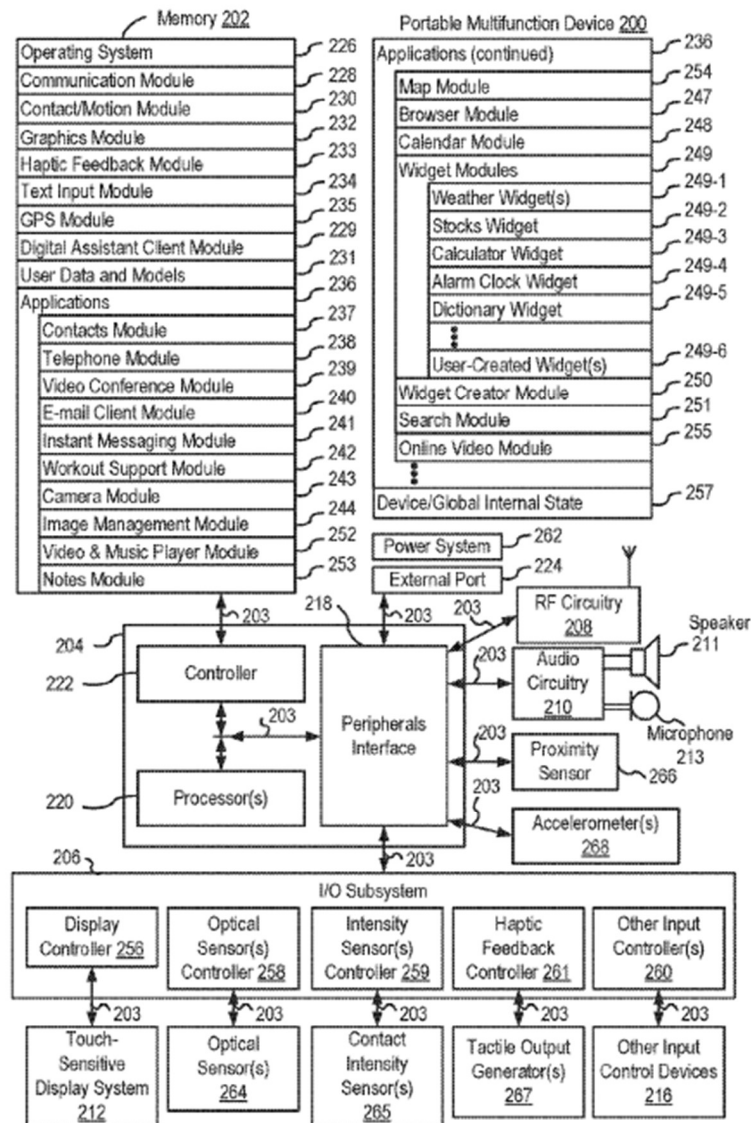


FIG. 2A

Raitio, FIG. 2A. Figure 8A below illustrates additional modules, including a “whispered speech determination module” of the digital assistant. Raitio, [0020], [0239]-[0240], FIG. 8A.

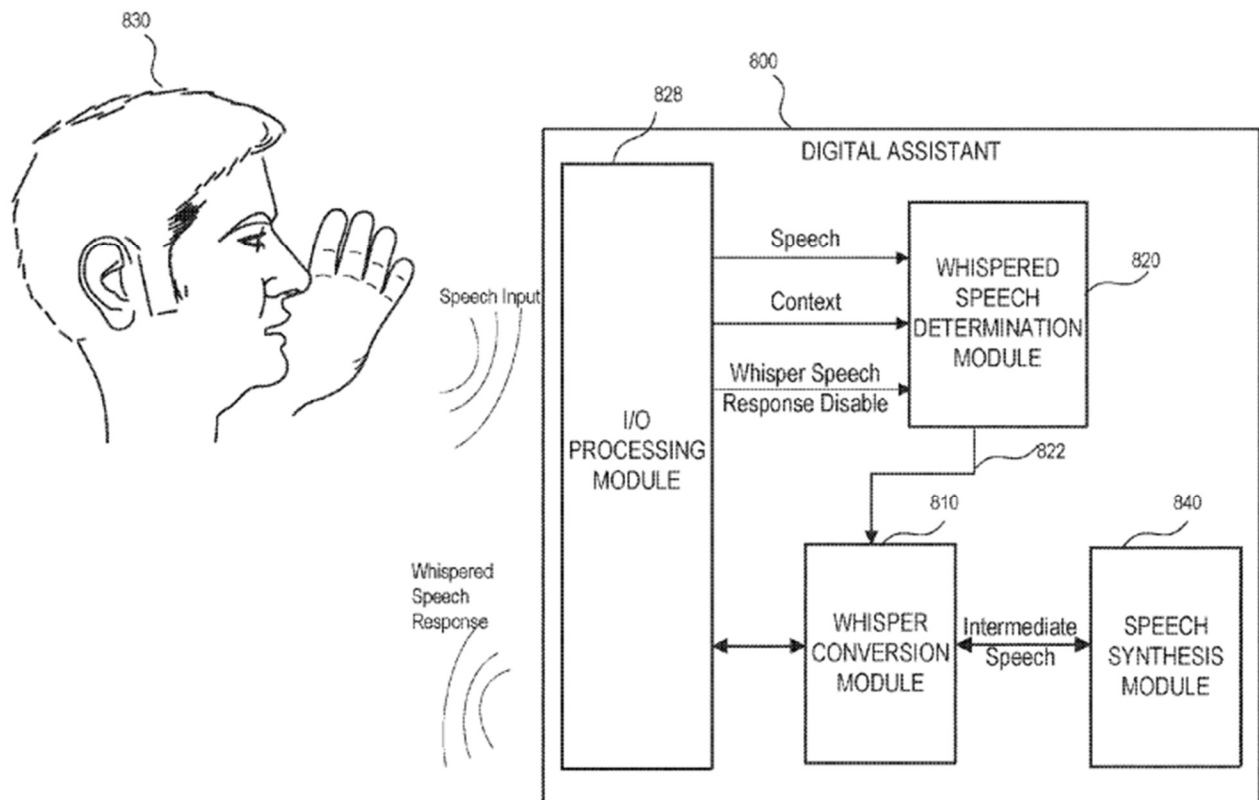


FIG. 8A

Raitio, FIG. 8A.

Like the '675 patent, Raitio's device “interprets natural language input ... to infer user intent, and performs actions based on the inferred user intent.” Raitio, [0032], [0033], [0036], FIG. 2A. In use, the device “accept[s] a user request” such

as a “natural language command, request, statement, narrative, and/or inquiry,” and “generat[es] output responses to the user in an audible (e.g., speech) and/or visual form” based on the determined “user’s intent.” Raitio, [0032], [0033], [0035], [0036], FIGS. 2A, 8A. The generated response may be whispered or non-whispered, depending on whether the device’s whispered speech determination module classifies the user’s speech input as whispered or non-whispered. Raitio, [0240], [0242]-[0250], [0253], FIGS. 8A, 9A-9B, claim 25. In this way, the device can account for the user’s need for discrete outputs and privacy. Raitio, [0028] (recognizing “various” reasons why it would be desirable to alter the speech output based on the whispered speech input determination, “such as avoiding disturbing others and protecting the user’s privacy”).

Raitio’s whispered speech determination module 820 “determine[s] whether the speech input includes a whispered speech input” using various “classification methods and/or spectral estimation methods.” Raitio, [0240], [0242]-[0250], FIGS. 2A, 8A, 9B. Like the voice classification module of the ’675 patent, the whispered speech determination module may use “one or more spectrum characteristics, such as the amplitude, the energy, the volume, the slope, or a combination thereof” of the user’s speech input, to determine whether whispered speech was input. Raitio, [0245] (further explaining that the spectrums may be compared to a threshold to classify the input as whispered or non-whispered); Ex.

1001, 9:60-10:16 (“In the present embodiment, the voice classifying unit 42 classifies the voice V1 to the first voice V1A when the voice V1 is determined to be a whisper as a result of analysis (determines as the predetermined voice).... The voice classifying unit 42 classifies the voice V1 either to the first voice V1A or the second voice V1B by using a peak frequency that is equal to or higher than a predetermined intensity in the spectrum as a feature value.”). Figure 9B below illustrates one of Raitio’s processes for determining whether the user’s speech input is a whisper or non-whisper. Raitio, [0285]-[0286], FIG. 9B.

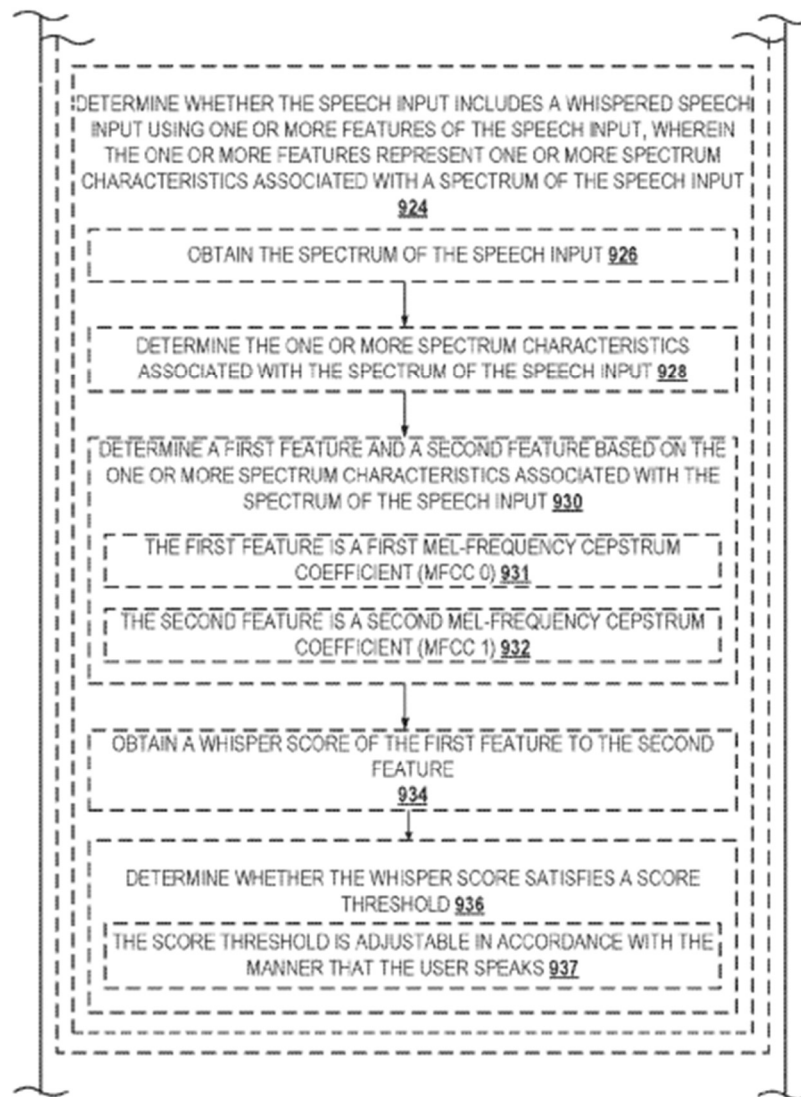


FIG. 9B

Raitio, FIG. 9B .

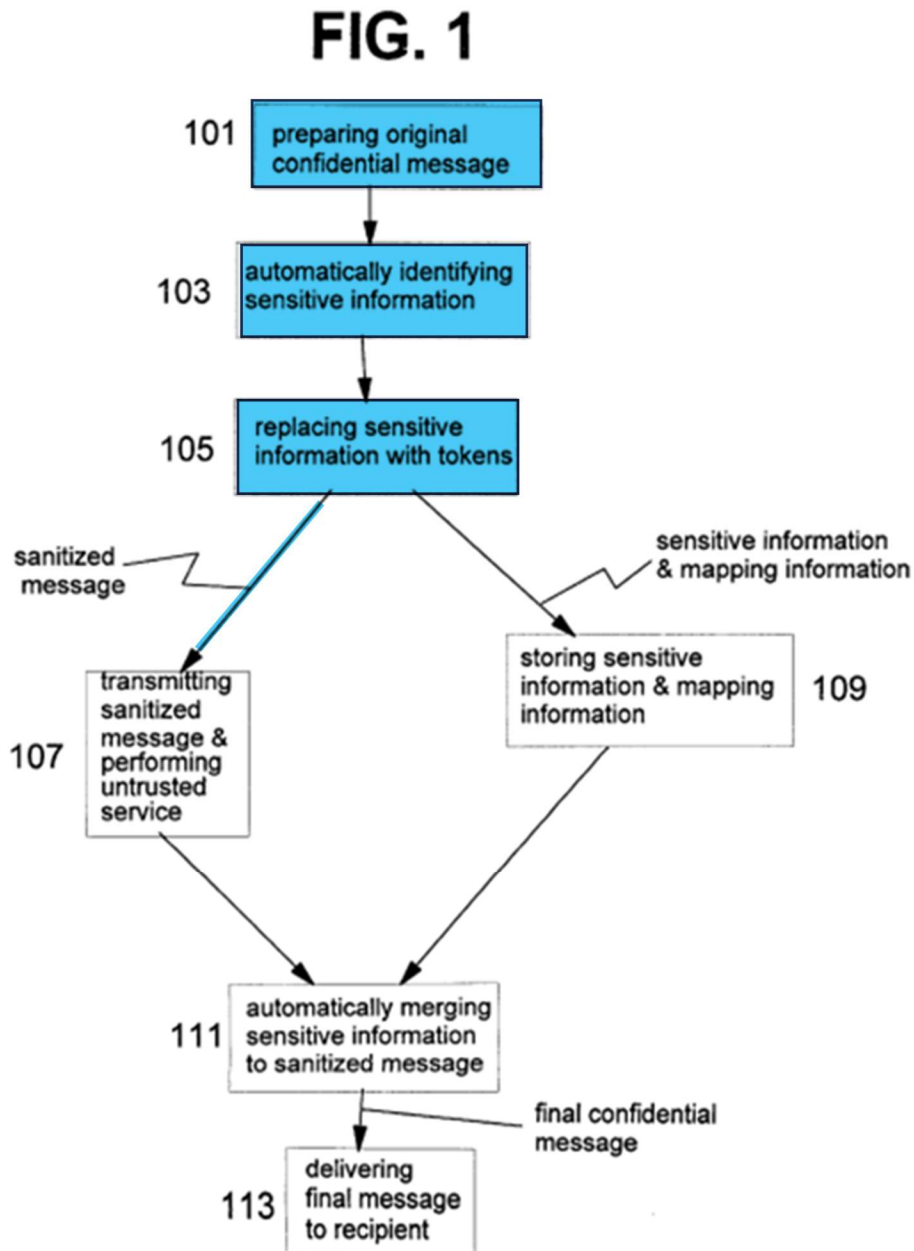
Raitio is analogous art to the '675 patent because it is in the same field of endeavor as the '675 patent: speech processing. *Compare* Raitio, Abstract, with Ex. 1001, Abstract; Lipoff, ¶71. Furthermore, Raitio is directed to the same problem as the '675 patent: analyzing voice input commands to generate

responses. *Compare* Raitio, Abstract, [0240]-[0241], FIG. 9A (steps 904, 906, 913), claims 2-3, *with* Ex. 1001, Abstract; Lipoff, ¶71.

B. Fahlman (Ex. 1007)

Fahlman, filed November 7, 1997, and issued September 28, 1999, qualifies as prior art to the '675 patent under at least 35 U.S.C. § 102(a)(1). Fahlman, cover (22), (45).

Fahlman teaches “a method and apparatus for transforming a message containing sensitive information” by replacing one or more “terms containing sensitive information” with a “standard token” to generate a “sanitized message.” Fahlman, Abstract, 1:7-12, 3:24-26, 3:35-53. Annotated Figure 1 below illustrates a process for sanitizing a confidential message by replacing one or more words in the confidential message.



Fahlman, FIG. 1 (annotated). Fahlman’s goal of sanitizing messages to preserve confidentiality comports with Raitio’s and the ’675 patent’s goal of generating discrete responses to accommodate a user’s need for privacy. *Compare* Fahlman, 1:66-2:3 (“An object of the present invention is to provide a method and an

apparatus for transforming a message by separating sensitive information from the message prior to an untrusted service, so that the confidentiality of the information is maintained.”), *with* Raitio, [0028] (“[T]echniques for determining that a whispered speech response is to be provided are desired for various purposes, such as avoiding disturbing others and protecting the user’s privacy.”), and Ex. 1001, 17:50-67 (“The output-content control device 1 determines that it is required to avoid the content being known by people other than the user H when the voice V1 classified into the first voice V1A is detected [R]eplacement of the word with the other word is performed, thereby enabling to make the content difficult to be understood by the people other than the user H.”). Thus, Fahlman is analogous art to the ’675 patent because it is directed to the same problem as the ’675 patent and is in the same field of endeavor as the ’675 patent: speech processing.

Compare Fahlman, 1:7-12, 1:66-2:3 *with* Ex. 1001, Abstract; Lipoff, ¶74.

C. Independent Claims

1. Claim 1

Raitio discloses or suggests every element of claim 1. While Raitio does not expressly disclose performing word replacement, as recited in element [1f], *infra*, Fahlman does.

a. [1preamble]: “An output-content control device comprising:”

Raitio discloses [1preamble] because it discloses a “portable multifunctional device” (an output-content control device) configured as a “digital assistant,” “virtual assistant,” “intelligent automated assistant,” or “automatic digital assistant” that “interprets natural language input ... to infer user intent, and performs actions based on the inferred user intent.” Raitio, [0032], [0033], [0036], FIG. 2A; Lipoff, ¶77. In use, the device “accept[s] a user request” such as a “natural language command, request, statement, narrative, and/or inquiry,” and “generat[es] output responses to the user in an audible (e.g., speech) and/or visual form” based on the determined “user’s intent.” Raitio, [0032], [0033], [0035], [0036], FIGS. 2A, 8A.

Figure 2A below “is a block diagram illustrating a portable multifunction device implementing the client-side portion of a digital assistant in accordance with some embodiments.” Raitio, [0009], FIG. 2A.

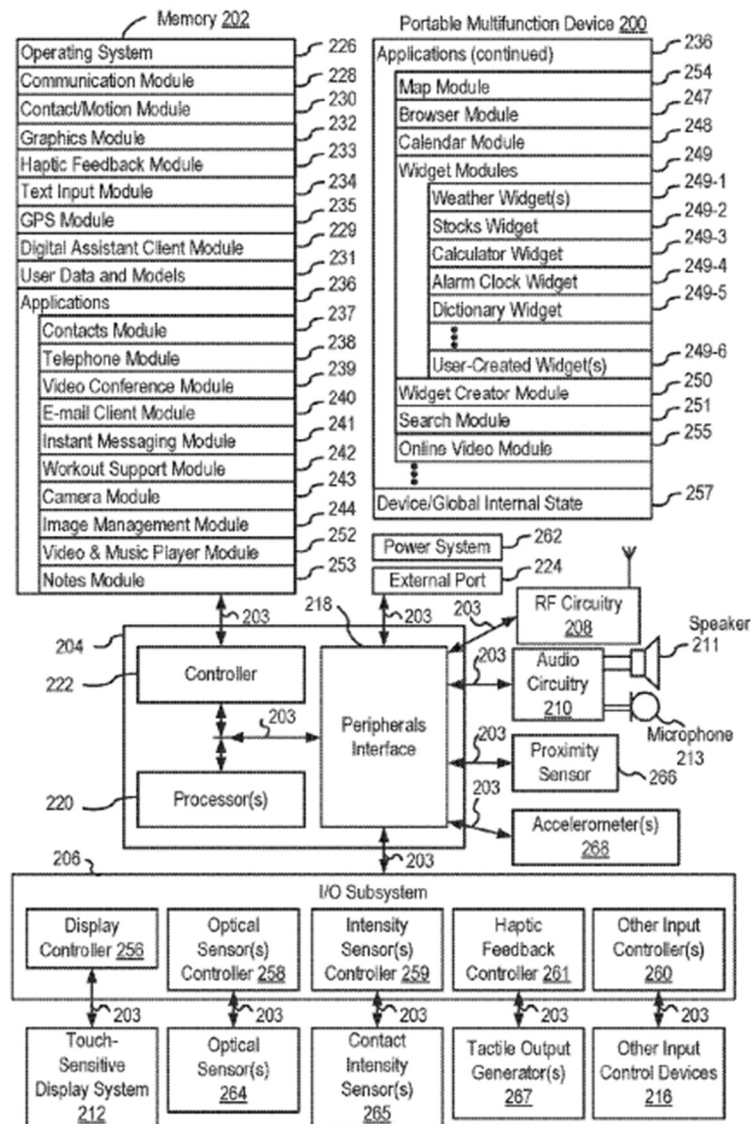


FIG. 2A

Raitio, FIG. 2A.

Figure 8A below illustrates a block diagram of the functions of the digital assistant. Raitio, [0020], [0239]-[0240], FIG. 8A.

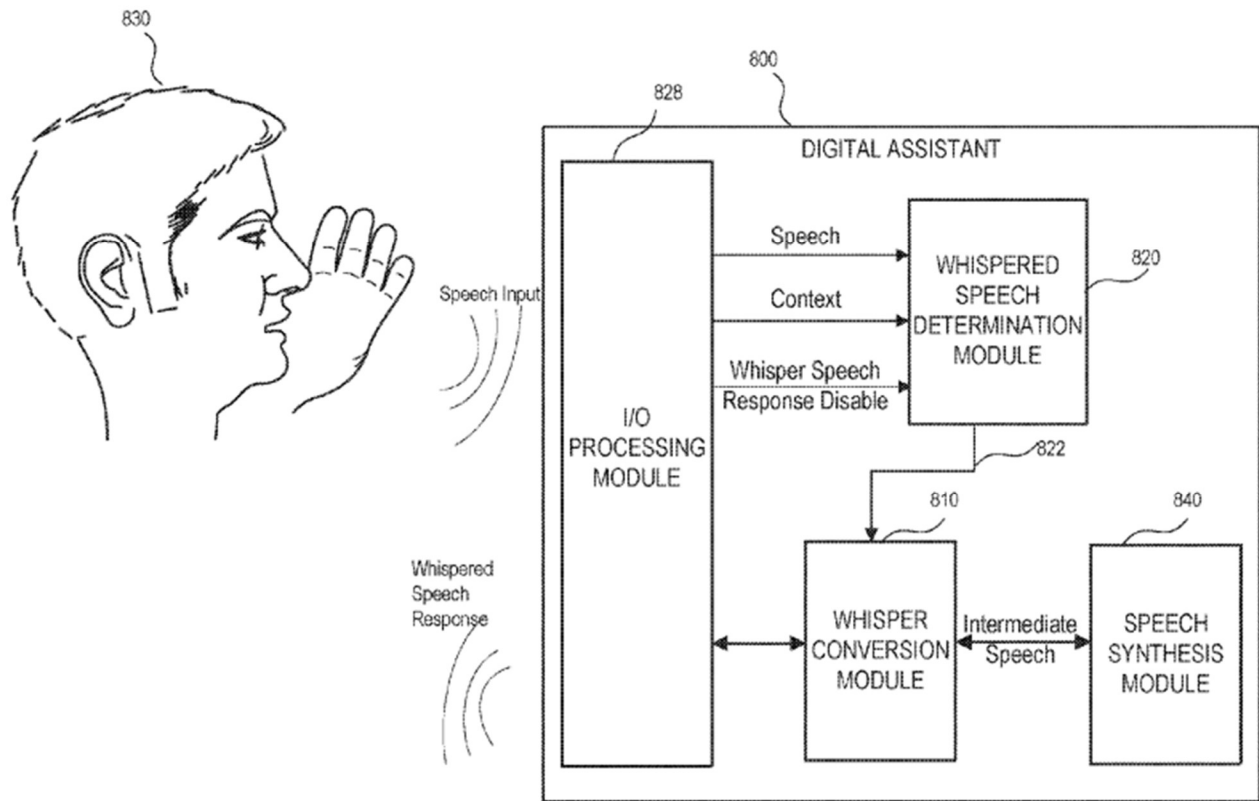


FIG. 8A

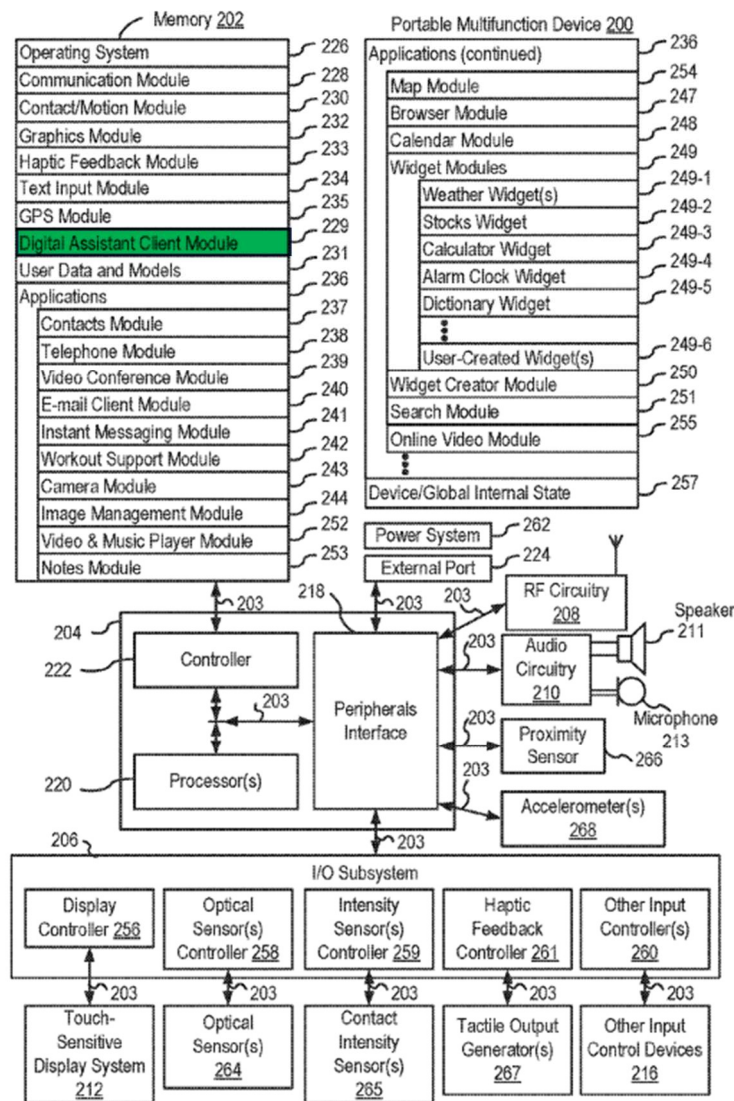
Raitio, FIG. 8A.

- b. [1a]: “a voice acquiring unit configured to acquire a voice spoken by a user;”

Raitio discloses [1a] because its “device”/“assistant” (output-content control device) includes a “digital assistant client module 229” (voice acquiring unit) “capable of accepting voice input (e.g., speech input),³ ... through various user

³ All emphases have been added, unless otherwise noted.

interfaces (e.g., microphone 213, ...) of portable multifunction device 200.” Raitio, [0077]; *see also* Raitio, [0207]-[0208] (describing a “digital assistant module 726” corresponding to digital assistant 700), [0239] (describing digital assistant 800 implementing the same digital assistant module); Lipoff, ¶79.



Raitio, FIG. 2A (annotated).

The device's memory (202/702) stores the module and its associated “applications (or sets of instructions).” Raitio, [0066], [0207], FIGS. 2A, 7A-7B. This matches the '675 patent, which explains that its voice acquiring module merely “acquires” a voice V1 that is “detected by” a “voice detecting unit” (also called an input unit), such as a “microphone.” Ex 1001, 3:27-33 (“The voice detecting unit 10 is a microphone”), 4:1-2.

- c. **[1b]: “a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”**

Raitio discloses [1b] because its “device”/“assistant” (output-content control device) includes a “**whispered speech determination module 820**” (voice classifying unit) that “can determine whether the speech input includes a whispered speech input” (a first voice) or “non-whispered speech input” (a second voice) using “any desired classification methods and/or spectral estimation methods,” including a “far field acoustic model” (proximity sensor) that “enables the detection of speech when the user speaks at an extended distance from [the] digital assistant.” Raitio, [0063], [0240], [0242]-[0250], [0285]-[0286], FIGS. 2A, 8A, 9B; Lipoff, ¶¶80-84.

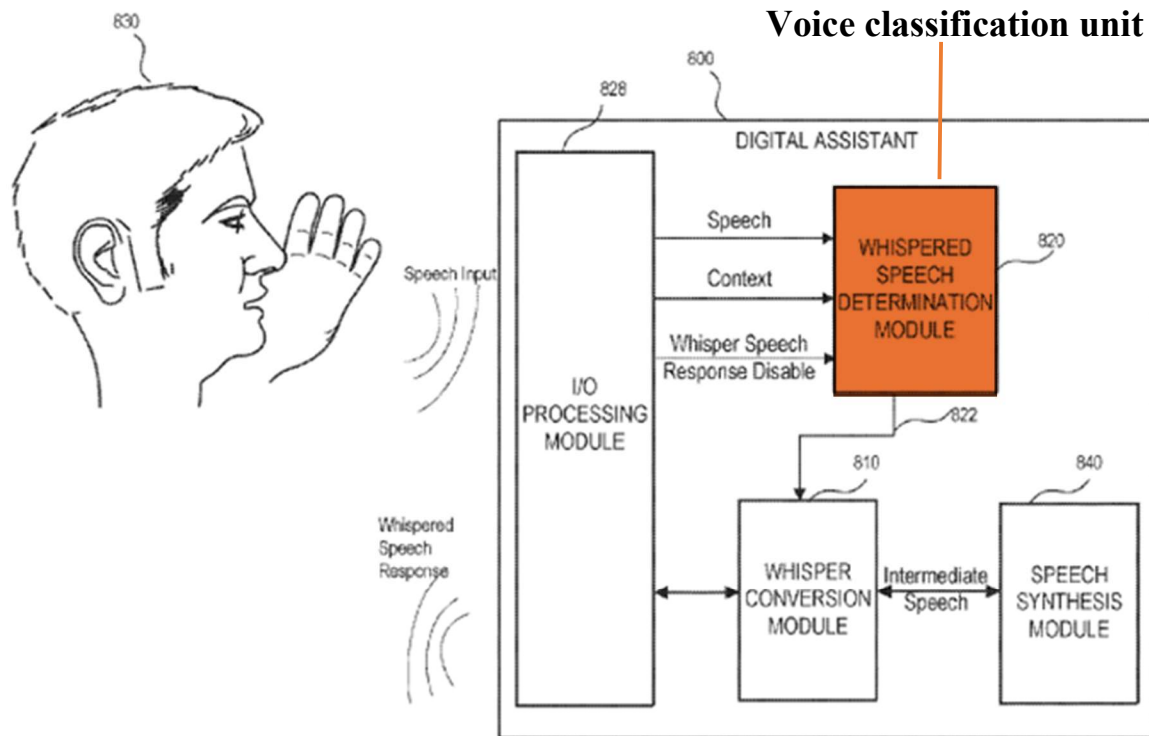


FIG. 8A

Raitio, FIG. 8A (annotated).

As Raitio explains, “[a] whispered speech input may be different from a non-whispered speech input in various aspects. For example, to generate a non-whispered speech input, vocal folds of user 830 vibrate and produce a periodic excitation signal” corresponding to a “non-whispered phonation (e.g., phonation in regular voice).” Raitio, [0243]. When the user “generate[s] a whispered speech input,” however, “the air passing through the glottis becomes turbulent, but not enough to start a vibration” such that an alternate “excitation signal” resulting in a

“whispered phonation” is generated. Raitio, [0243]. These excitation signals corresponding to “whispered” and “non-whispered” speech inputs “have differing spectrums” that the module analyzes to classify the input as “whispered” or “non-whispered” speech. Raitio, [0243]-[0244]. For example, “whispered speech determination module 820 can make such determination based on one or more spectrum characteristics, such as the amplitude, the energy, the volume, the slope, or a combination thereof.” Raitio, [0245] (further explaining that the spectrums may be compared to a threshold to classify the input as whispered or non-whispered).

Raitio’s whispered and non-whispered speech inputs are first and second voices within the meaning of the ’675 patent. Ex. 1001, 9:65-10:4 (“[T]he voice classifying unit 42 classifies the voice V1 to the first voice V1A when the voice V1 is determined to be a whisper The voice classifying unit 42 classifies the voice V1 to the second voice V1B when the voice V1 is determined not to be a whisper.”); Lipoff, ¶81.

Raitio further explains that “[w]hen digital assistant 800 is in the far field, the user’s speech may appear to be soft, and the spectrum of the speech input that digital assistant 800 receives may have a low volume, amplitude, and/or energy.” Raitio, [0249]. That is, “a whispered speech input may have characteristics (e.g., lower volume, energy, and/or amplitude) similar to those of

the speech input that digital assistant 800 receives from a far-field user speech.”

Raitio, [0249]. Accordingly, the whispered speech determination module may implement a “far-field acoustic model” that “enables the detection of speech when the user speaks at *an extended distance* from digital assistant 800 (e.g., at a distance beyond a threshold distance).” Raitio, [0249]. A POSITA would have understood that because the far-field acoustic model uses a distance-based threshold to determine user position and detect whispering, it calculates a distance between the user and the device. Lipoff, ¶82. This model (a proximity sensor) “can increase the accuracy of whispered speech input detection.” Raitio, [0249]. Figure 9B below illustrates a process for classifying an input as a first voice (whispered speech). Raitio, [0285]-[0286], FIG. 9B.

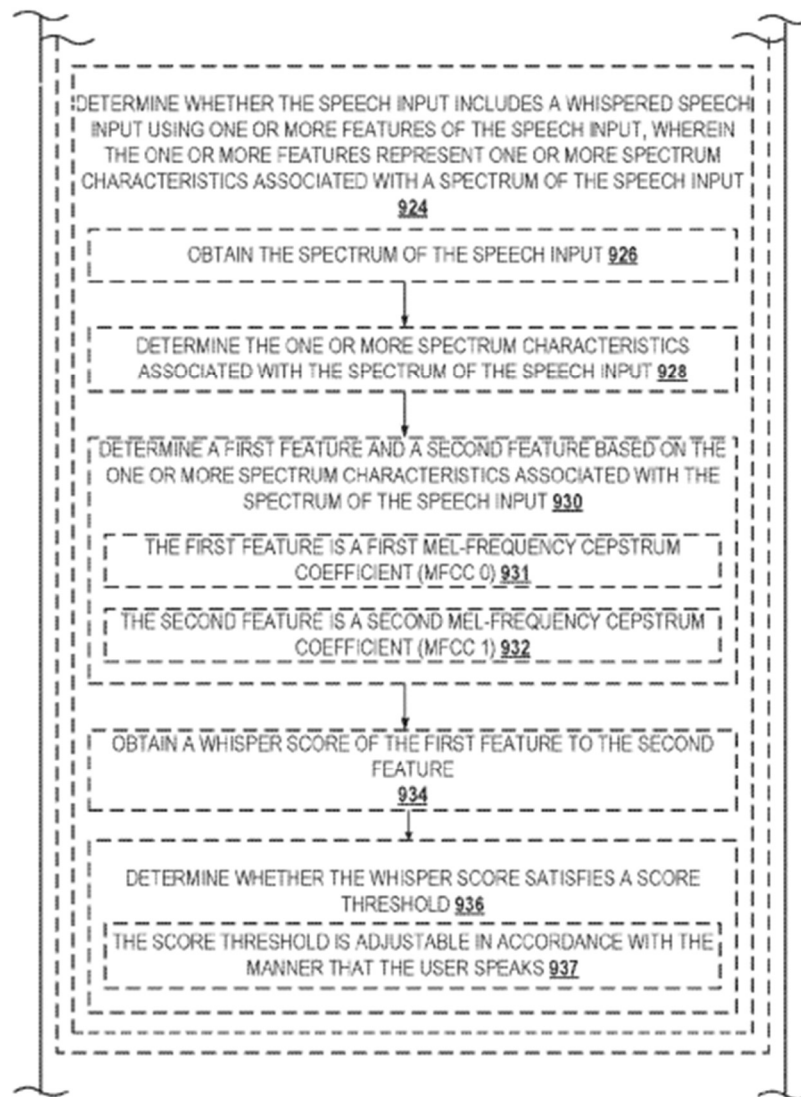


FIG. 9B

Raitio, FIG. 9B

Raitio's disclosure is consistent with the '675 patent, which does not limit its "proximity sensor" to a particular configuration, but explains that "a distance between the user H and the output-content control device 1 is calculated from a detection result of the proximity sensor, and the distance can be used as a feature

value to perform the classification to the first voice V1A and the second voice V1B.” Ex. 1001, 10:32-38; Lipoff, ¶83.

Furthermore, Raitio discloses that its device may include “one or more proximity sensors” such as those taught in U.S. Application No. 11/586,862, “Automated Response To And Sensing Of User Activity In Portable Devices” (Ex. 1011), and U.S. Application No. 11/638,251, “Methods And Systems For Automatic Configuration Of Peripherals” (Ex. 1012), both of which are incorporated by reference in their entirety. Raitio, [0063]. As Mr. Lipoff explains, a POSITA would have understood that such sensors, which are configured to determine user proximity with respect to the device, and to relay this determination to Raitio’s digital assistant client module, would have similarly calculated a distance between the user and the output-content control device to enable voice classification, as was known in the art. Lipoff, ¶84 (citing Ocampo, [0058], [0060]-[0062], [0065] (explaining that “touch, optical, or capacitive sensors may be used to determine whether a user is touching the user device or is in close proximity of the user device,” and using the outputs of such sensors to “classify audio signals received from a user” as a “whispering, shouting, happy, sad, or excited” voice); Ex. 1011, [0087], [0090]; Ex. 1012, [0094]-[0095]).

Accordingly, Raitio discloses a voice classifying unit configured to calculate, using its far-field model, a distance between the user and the digital

assistant (output-content control device) to classify the voice into either a whispered voice (first voice) or a non-whispered voice (second voice) based on the calculated distance. Raitio, [0063], [0240]-[0250], [0285]-[0286], FIGS. 2A, 8A, 9B; Lipoff, ¶¶80-85.

- d. [1c]: “an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;”

Raitio discloses [1c] because its “device”/“assistant” (output-content control device) includes “[o]ne or more processing modules” such as a “[natural language processing module](#)” (an intention analyzing unit) that “can utilize data and models ... to process speech input and *determine the user’s intent based on natural language input.*” Raitio, [0035], [0078], [0079], [0207]-[0208], [0210], [0216], [0239], [0256], FIG. 7B; Lipoff, ¶86; *see also* Raitio, [0032] (digital assistant “refer[s] to any information processing system that interprets natural language input in spoken and/or textual form to infer user intent, and performs actions based on the inferred user intent”).

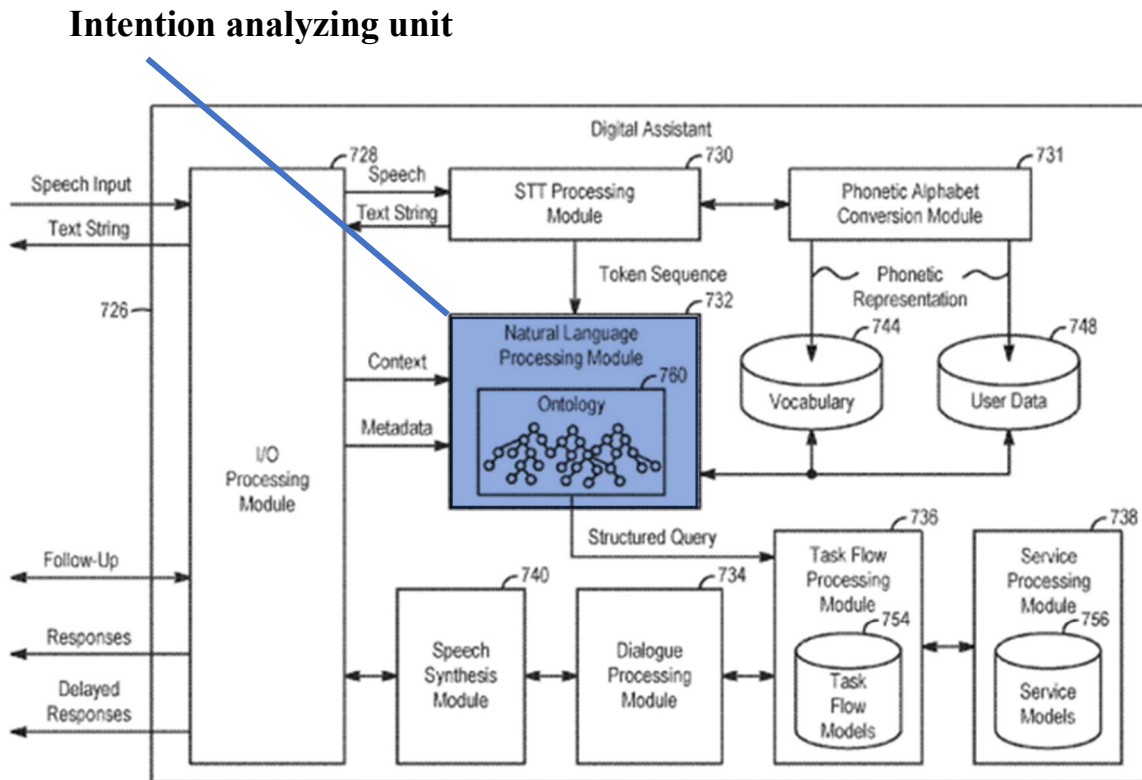


FIG. 7B

Raitio, FIG. 7B (annotated).

As Raitio explains, “the digital assistant can perform at least some of the following: converting speech input into text; *identifying a user’s intent expressed in a natural language input received from the user; actively eliciting and obtaining information needed to fully infer the user’s intent* (e.g., by disambiguating words, games, intentions, etc.); determining the task flow for fulfilling the inferred intent; and executing the task flow to fulfill the inferred intent.” Raitio, [0208], [0256].

For example, a “[n]atural language processing module 732 (‘natural language processor’) of the digital assistant can *take the sequence of words* or

tokens ('token sequence') generated by [speech-to-text] STT processing module 730, *and attempt to associate the token sequence with one or more 'actionable intents' recognized by the digital assistant.*" Raitio, [0216],

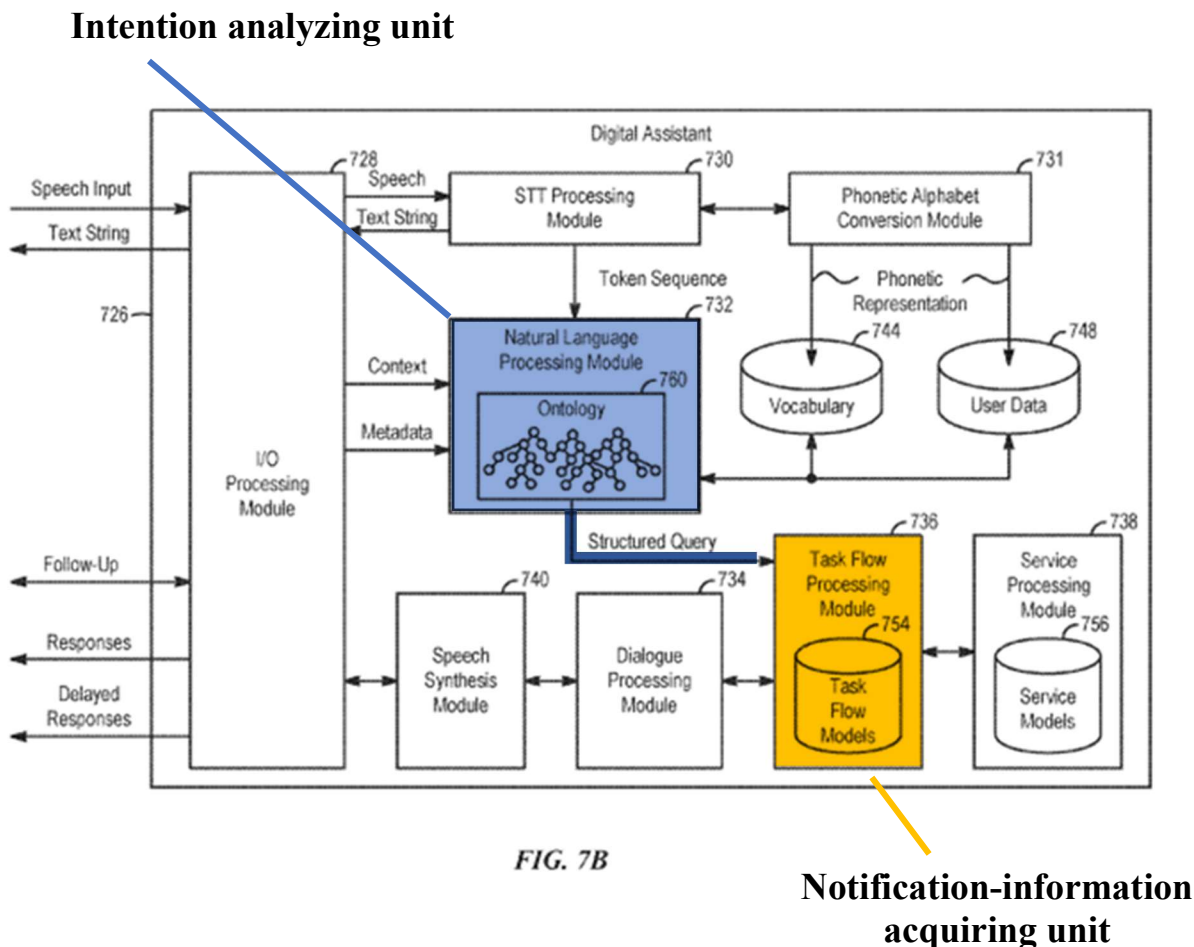
FIG. 7B. As illustrated in Figure 7B above, the user's speech input is received by the STT module, which uses automatic speech recognition (ASR) techniques to process the input and passes the "recognition results containing a text string (e.g., words, or sequence of words, or sequence of tokens) ... to natural language processing module 732 for *intent deduction*." Raitio, [0210]. The natural language processing module then processes the tokens (using, e.g., "ontology") to identify "actionable intents" that "represent a task that can be performed by the digital assistant" in response to the user's request, and "select[s] one of the actionable intents as the task that the user intended the digital assistant to perform." Raitio, [0216], [0218], [0226], [0228]; *see also* Raitio, [0219]-[0225], [0229] (describing methods for deducing actionable intents), [0239] (explaining that digital assistant 800 also implements the natural language processing module), [0256] (explaining that digital assistant 800 "actively elicit[s] and obtain[s] information needed to fully infer the user's intent (e.g., by disambiguating words, games, intentions, etc.)").

- e. **[1d]: “a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and”**

Raitio discloses [1d] because its device/digital assistant includes a “task flow processing module” (notification-information acquiring unit) configured to “obtain[] additional information” (acquire notification information including content information) based on the user’s determined intent. Raitio, [0229], [0230], [0232], [0235], FIG. 7B; Lipoff, ¶90. The acquired information is subsequently presented to the user. Raitio, [0229], [0230], [0232], [0235], FIG. 7B; Lipoff, ¶90.

As explained for [1c], *supra*, Raitio’s natural language processing module processes “words, word strings, or sequence of tokens” to identify “actionable intents” that “represent a task that can be performed by the digital assistant” in response to the user’s request, and “select[s] one of the actionable intents as the task that the user intended the digital assistant to perform.” Raitio, [0210], [0216], [0218], [0226], [0228]. The natural language processing module may also “generate a structured query to represent the identified actionable intent” and use a task flow processing module to “complete[] the structured query for an actionable intent.” Raitio, [0229], [0232]. The task flow module obtains “additional information” (content information) from the user or various stored “task flow” models and populates the structured query with content that is responsive to the

user's inferred intent so that it may be provided to the user in a response (notified to the user based on the intention information). Raitio, [0230]. Digital assistant 800 illustrated in Figure 8A also implements the task flow processing module. Raitio, [0239], [0256] (explaining that digital assistant 800 “determin[es] the task flow for fulfilling the inferred intent” and “execut[es] the task flow to fulfill the inferred intent”), FIG. 8A.



Raitio, FIG. 7B (annotated).

For example, when the speech input includes “a request to perform a task” such as “Make me a dinner reservation at a sushi place at 7,” the natural language processing module may infer that the user desires a “restaurant reservation,” and the task flow module may obtain content information relating to the desired party size and restaurant to fulfill the request. Raitio, [0229], [0232], [0240]-[0241], [0254]; *see also* Raitio, [0231] (“[T]ask flow processing module 736 invokes dialogue flow processing module 734 to determine the ‘party size’ and ‘date’ information for the structured query associated with the domain ‘restaurant reservation’”). The acquired information (e.g., “party size=5” and “restaurant=ABC Café”) completes the structured query so that it may be notified to the user via, e.g., a “calendar entry ... in the user’s calendar.” Raitio, [0232].

When “speech input” includes an “informational request” such as “What time is it in Paris now?” the natural language processing module may infer that the user desires to know the current time in Paris, and the task flow module may retrieve information (content) from the task flow model to produce an “informational answer” such as “It is 2 p.m. in Paris right now.” Raitio, [0240], [0257]; *see also* Raitio, [0216] (“The associated task flow can be a series of programmed actions and steps that the digital assistant takes in order to perform the task.”), [0230], [0233], [0240], [0033] (describing other informational requests and answers).

- f. **[1e]: “an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information,”**

Raitio discloses [1e] because its device/digital assistant contains software configured to “generate a response” such as “a dialogue response” that “at least partially fulfills the user’s intent” using the content (notification information) retrieved by the task flow processing module (an output-content generating unit). Raitio, [0235]-[0236]; Lipoff, ¶¶93-94. The “generated dialogue response can be in the form of a text string” (outputted sentence data) and “sent to speech synthesis module 740 (e.g., speech synthesizer)” that “convert[s] the text string to an audible speech output.” Raitio, [0235]-[0236]; *see also* Raitio, [0239], [0256], FIG. 8A (describing and illustrating digital assistant 800 implements the speech synthesis module).

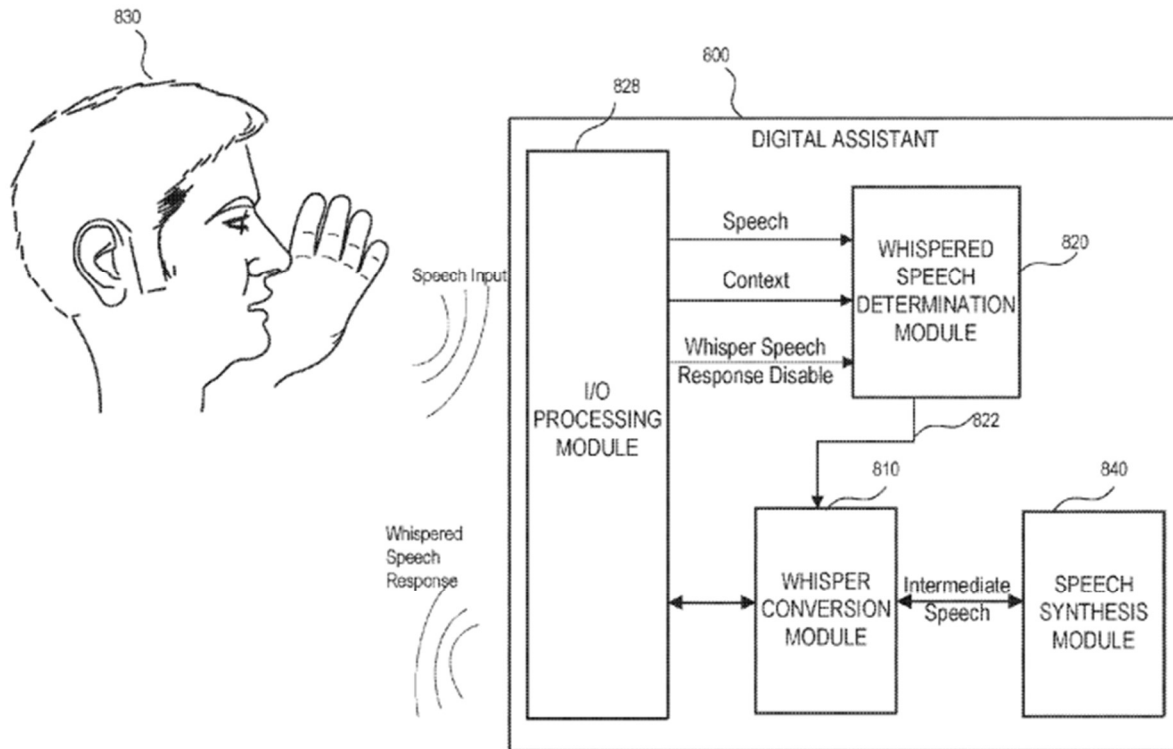


FIG. 8A

Raitio, FIG. 8A.

The generated output sentence is also “based on the notification information”—mapped to the additional information obtained by Raitio in [1d]. Lipoff ¶94. As explained in [1d], the acquired notification information to be notified to the user is based on the intention information. For example, when “speech input” includes an “informational request” such as “What time is it in Paris now?” the natural language processing module may infer that the user desires to know the current time in Paris, and the task flow module may retrieve information, i.e., the current time (notification information including content

information), from the task flow model to produce an “informational answer” such as “It is 2 p.m. in Paris right now.” Raitio, [0240], [0257]; *see also* Raitio, [0216] (“The associated task flow can be a series of programmed actions and steps that the digital assistant takes in order to perform the task.”), [0033] (“A satisfactory response to the user request can be a provision of the requested informational answer, a performance of the requested task, or a combination of the two. For example, a user can ask the digital assistant a question, such as “Where am I right now?” *Based on the user’s current location, the digital assistant can answer, “You are in Central Park near the west gate.”*).

Raitio further explains “the generated response” (output based on notification information) “can be sent to speech synthesis module 740 (e.g., speech synthesizer) ...[,]processed to synthesize the dialogue response in speech form,” and output as “audible speech.” Raitio, [0235], [0236] (“Speech synthesis module 740 synthesizes speech outputs based on text provided by the digital assistant ... Speech synthesis module 740 can convert the text string to an audible speech output”). Thus, the generated output sentence data (e.g., “You are in Central Park near the west gate” or “It is 2 p.m. in Paris right now,”) to be output to the user is based on the notification information (the user’s current location or time in Paris). Raitio, [0235]-[0236], [0240], [0247]; Lipoff ¶94.

Fahlman similarly discloses a system and method for generating an output message (sentence), as explained in further detail in [1f], *infra*. Fahlman, 3:35-60; Lipoff, ¶94.

- g. **[1f]: “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word, and to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”**

Raitio-Fahlman renders obvious [1f]. Lipoff, ¶¶95-105.

Raitio alone at least suggests [1f] because Raitio teaches modifying the generated response (content information of the notification information) when the voice is determined to be “whispered speech” (the first voice) versus “non-whispered speech” (the second voice). Raitio, [0242]-[0243], [0254], [0257]-[0258]; Lipoff, ¶96.

As explained for [1e], *supra*, Raitio discloses “generat[ing] a response” such as “a dialogue response” that “at least partially fulfills the user’s intent” using the content (notification information) retrieved by the task flow processing module. Raitio, [0235]; Lipoff, ¶97. And as explained for [1b], *supra*, “[w]hispered speech determination module 820 can determine whether the speech input includes a

whispered speech input based on one or more differences between a whispered speech input and a non-whispered speech input.” Raitio, [0243], [0249]; Lipoff, ¶97.

Raitio further discloses altering the generated response based on its determination that the speech input is a whisper or non-whisper (i.e., a first voice or second voice)—that is, “if whispered speech determination module 820 determines that the speech input includes a whispered speech input and/or that the context data indicates a whispered speech response is expected.” Raitio, [0254]. For example, if the user asks, “What time is it in Paris now?” in a non-whispered voice, the digital assistant 800 may answer, “It is 2 p.m. in Paris right now.” Raitio, [0240]. However, if the user “whisper[s], ‘What is the time now in Paris?’” the digital assistant may “generate and provide a *whispered* speech response such as ‘The current time in Paris is 2 p.m.’” Raitio, [0254]. Raitio further recognizes that there are “various” reasons why it would be desirable to alter the speech output based on the whispered speech input determination, “such as avoiding disturbing others and protecting the user’s privacy.” Raitio, [0028]. Altering the response, Raitio explains, “can allow the user to use the digital assistant under certain circumstances without having to change the user’s current physical location (e.g., step outside to use the digital assistant).” Raitio, [0028].

Raitio does not expressly disclose performing word replacement based on a determination that the voice is whispered speech (low amplitude, volume, etc.), but Fahlman does.

Fahlman teaches “a method and apparatus for transforming a message containing sensitive information” by replacing one or more “terms containing sensitive information” with a “standard token” to generate a “sanitized message.” Fahlman, Abstract, 1:7-12, 3:24-26, 3:35-53. “For example, the sensitive term ‘Mr. Johnson’ is replaced by the standard token <person-1>, and the term ‘Jul. 1, 1997’ is replaced by <date-1>.” Fahlman, 3:48-53. As Fahlman explains, “[t]he sanitized message,” which may be text or voice, “is then transmitted with a low level of security since the sanitized message only contain[s] standard tokens, not sensitive information.” Fahlman, 3:29-33, 3:54-60.

i. Motivation to Combine Raitio and Fahlman

A POSITA would have appreciated that Fahlman’s word replacement would have been useful for generating responses that obviate the “privacy” concerns of Raitio’s user. Raitio, [0028]; Lipoff, ¶100. For example, a POSITA would have appreciated that in addition to whispering a generated response (as described in above), sanitizing the response to remove any private (sensitive) information would have further enhanced the user’s experience by reducing the likelihood that others in the user’s vicinity would also understand the response. Lipoff, ¶100.

Thus, a POSITA would have been motivated to implement Fahlman's word replacement algorithms in Raitio's system to remove any private or sensitive information from the response that the user may not want others in his/her proximity to hear. Lipoff, ¶¶100-101.

ii. Expectation of Success

A POSITA would have had a reasonable expectation of success in implementing Fahlman's algorithm because it would have required only routine skill to do so. Lipoff, ¶¶102-105.

As Raitio explains, "[e]ach of the above-identified modules and applications," including its dialogue generation processes, "corresponds to a set of executable instructions for performing one or more functions described above and the methods described in this application (e.g., the computer-implemented methods and other information processing methods described herein)." Raitio, [0033], [0039], [0121].

Fahlman similarly discloses that "a suitable computer program algorithm" may be used to generate the sanitized message. Fahlman, 3:35-53. For example, an algorithm may be used to identify the sensitive terms contained in the message using, e.g., a "defined list of sensitive terms, such as known places, known names, dollar (\$) signs, etc." Fahlman, 3:35-47, 6:29-34. And "another subroutine or various combination[s] of algorithms" may be used "for replacing the identified

sensitive terms with standard tokens, thereby creating a sanitized message.”

Fahlman, 6:35-38.

Thus, a POSITA would have had a reasonable expectation of success in implementing Fahlman’s word replacement algorithms in Raitio’s assistant because it would have merely required the integration of computer algorithms and routine skill to do so. Lipoff, ¶¶102-105.

2. Claim 6

a. [6preamble]: “An output-content control method comprising:”

Raitio discloses [6preamble] because it discloses “processes for operating a digital assistant,” including “receiving a speech input from a user” and “determining, based on the speech input, that a whispered speech response is to be provided,” and, “upon determining that a whispered speech response is to be provided, generating a whispered speech response, and providing the whispered speech response to the user.” Raitio, [0005], FIGS. 9A, 10; Lipoff, ¶123.



Raitio, FIG. 9A.

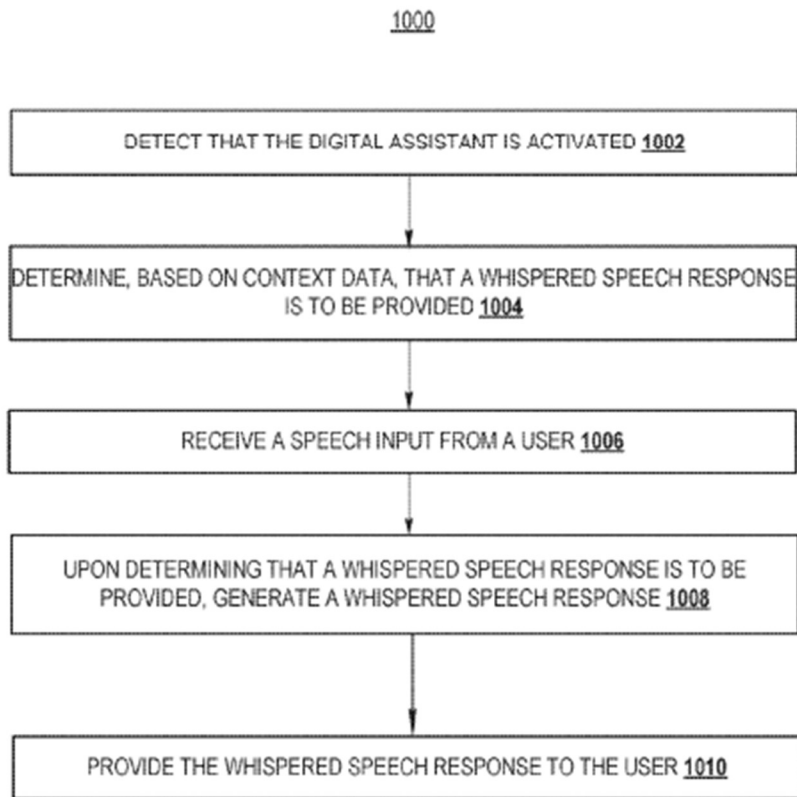


FIG. 10

Raitio, FIG. 10.

b. [6a]: “acquiring a voice spoken by a user;”

Raitio discloses [6a] for reasons discussed for [1a], *supra*. Lipoff, ¶124.

c. [6b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”

Raitio discloses [6b] for reasons discussed for [1b], *supra*. Lipoff, ¶125.

- d. **[6c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”**

Raitio discloses [6c] for reasons discussed for [1c], *supra*. Lipoff, ¶126.

- e. **[6d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”**

Raitio discloses [6d] for reasons discussed for [1d], *supra*. Lipoff, ¶127.

- f. **[6e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and”**

Raitio-Fahlman renders obvious [6e] for reasons discussed for [1e]-[1f], *supra*. Lipoff, ¶128.

- g. **[6f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”**

Raitio-Fahlman renders obvious [6f] for reasons discussed for [1e]-[1f], *supra*. Lipoff, ¶129.

3. Claim 7

- a. **[7preamble]: “A non-transitory storage medium that stores an output-content control program that causes a computer to execute.”**

Raitio discloses [7preamble] because its digital assistant includes “memory” that stores a “digital assistant module” for causing the digital assistant to perform

the functions recited in [7a]-[7f], *infra*. Raitio, [0006], [0202], [0207]-[0208], FIG. 7A; Lipoff, ¶130.

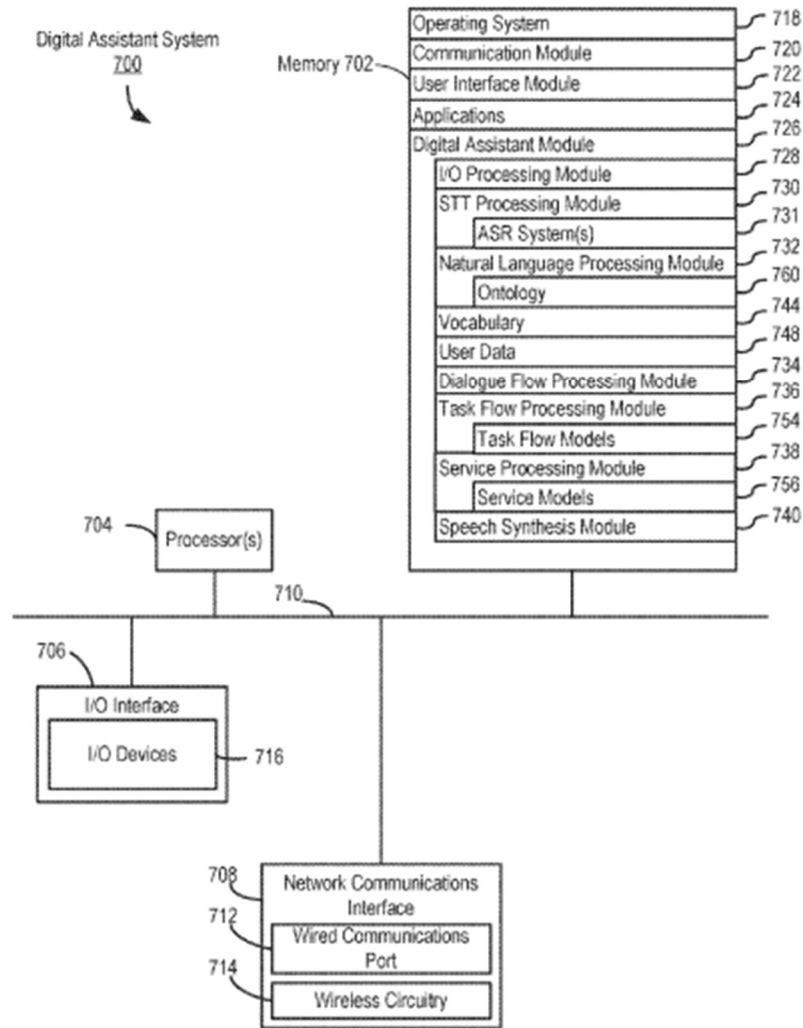


FIG. 7A

Raitio, FIG. 7A.

b. [7a]: “acquiring a voice spoken by a user;”

Raitio discloses [7a] for reasons discussed for [1a], *supra*. Lipoff, ¶131.

- c. **[7b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”**

Raitio discloses [7b] for reasons discussed for [1b], *supra*. Lipoff, ¶132.

- d. **[7c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”**

Raitio discloses [7c] for reasons discussed for [1c], *supra*. Lipoff, ¶133.

- e. **[7d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”**

Raitio discloses [7d] for reasons discussed for [1d], *supra*. Lipoff, ¶134.

- f. **[7e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and”**

Raitio-Fahlman renders obvious [7e] for reasons discussed for [1e]-[1f], *supra*. Lipoff, ¶135.

- g. **[7f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”**

Raitio-Fahlman renders obvious [7f] for reasons discussed for [1e]-[1f], *supra*. Lipoff, ¶136.

D. Dependent Claims

1. **Claim 2: “The output-content control device according to claim 1, wherein the notification information includes type information indicating a type of the content information, and the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be [re]placed with the other word based on a word included in the content information and the type information that is associated with the content information.”**

Raitio-Fahlman renders obvious claim 2 as explained below. Lipoff, ¶¶106-113.

“the notification information includes type information indicating a type of the content information”

As explained for [1d], *supra*, Raitio discloses acquiring notification information to be notified to the user, and that the notification information includes content information, i.e., the words to be conveyed to the user based on the user’s query. Raitio, [0229]-[0233]. The content information in the output also corresponds to a type of information. For example, the word/content “Paris” in the notification information “It is 2 p.m. in Paris right now” corresponds to the type of information *place*. Raitio, [0240], [0257]. This matches the ’675 patent, which explains that “type information” describes a category of information such as “person,” “date,” or place, while “content information” describes specific content

in the category, e.g., “Mr. Yamada,” “March 20, 2020,” or “Tokyo station.” Ex. 1001, 5:31-53, 8:21-44, 16:17-21, FIGS. 4-5.

In the combination presented above, Fahlman also discloses that its notification information includes not only the words to be notified to the user (content information) but categories (type information) indicating the type of content. For example, the content information “Mr. C. Williams” corresponds to the category (type information) “names of people or organizations,” while the content information “Jul. 7, 1997” corresponds to the category “dates & other time description.” Fahlman, 4:31-37, 5:41-6:16, FIG. 3. Categorizing the content information by type enables the device to identify which words to output and which words to replace in the sanitized message. Fahlman, 4:20-46; *see also* Raitio, [0232] (describing a “structured query” that separates the words in user queries and responses into distinct categories to facilitate query fulfillment).

Thus, to the extent Raitio does not expressly disclose that its notification information includes type information, a POSITA would have been motivated to modify the notification information such that it includes type information indicating a type of the content information, to facilitate user query fulfillment by allowing the device to easily identify which words to output in response to the user’s query. Fahlman, 4:20-46; Raitio, [0232]; Lipoff, ¶¶ 107-109. A POSITA

would have had a reasonable expectation of success in making such a modification for reasons articulated in [1f], *supra*. Lipoff, ¶¶107-109.

“the output-content generating unit further configured to select, on generating the first output sentence, at least one word to be [re]placed with the other word based on a word included in the content information and the type information that is associated with the content information”

As explained for [1f], *supra*, Fahlman teaches and renders obvious replacing one or more “terms containing sensitive information” with a “standard token” to generate a “sanitized message.” Fahlman, Abstract, 1:7-12, 3:24-26, 3:35-53.

Fahlman further teaches that the sensitive term (selected word to be replaced) is based on a word included in the content information and associated type information.

For example, when the notification information is “I have interviewed Mr. C. Williams and believe that we should hire him at an annual salary of \$80,000 commencing on Jul. 7, 1997,” the subroutine replaces the sensitive term (at least one word) “Mr. C. Williams” with the standard token (another word) “<person-3>” based on its classification as a sensitive term (type information) and its inclusion in the content information. Fahlman, 5:41-6:16, FIGS. 3-4.

Thus, Raitio’s output-content generating unit, as modified by Fahlman, teaches the output-content generating unit further configured to select, on

generating the first output sentence, at least one word to be replaced with the other word based on a word included in the content information and the type of information that is associated with the content information. Lipoff, ¶¶110-113.

2. **Claim 3: “The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.”**

Raitio-Fahlman presented above renders claim 3 obvious. Lipoff, ¶¶114-116.

Fahlman further explains that its computer program/subroutine (output-content generating unit) is configured to read “a defined list of sensitive terms” and corresponding “standard tokens” (relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word) and to replace the sensitive term (the first word matching a word in the notification information) with the standard token (the second word). Fahlman, 3:35-53, FIGS. 3-4.

For example, when the notification information is “I have interviewed Mr. C. Williams and believe that we should hire him at an annual salary of \$80,000 commencing on Jul. 7, 1997,” the subroutine identifies the sensitive term “Mr. C.

Williams” from its “defined list of sensitive terms” (first word that is predetermined) and replaces the sensitive term (a word included in the notification information/the first word) “Mr. C. Williams” with the standard token (second word associated with the first word) “<person-3>.” Fahlman, 4:44-47, 5:41-6:16, FIGS. 3-4.

3. **Claim 4: “The output-content control device according to claim 3, wherein the relationship information further includes information of type information indicating a type of the first word, and the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word.”**

Raitio-Fahlman renders claim 4 obvious because Fahlman further discloses that the “defined list of sensitive terms” and corresponding “standard tokens” (relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word) also include category information indicating a type of the sensitive term (type information indicating a type of the first word). Fahlman, 4:31-37, 4:44-47, 5:41-6:16, FIGS. 3-4; Lipoff, ¶¶117-119.

For example, the relationship information includes the categories “names of people or organizations.” Fahlman, 4:31-37, FIG. 3. Fahlman’s computer

program/subroutine (output-content generating unit) is also configured to select a sensitive term for replacement with a standard token first (a word included in the notification information as the word to be replaced with another word) when the sensitive term belongs to the same category (type) as the term in the notification information, and the word in the notification information matches the sensitive term.

For example, when the notification information is “I have interviewed Mr. C. Williams and believe that we should hire him at an annual salary of \$80,000 commencing on Jul. 7, 1997,” the subroutine replaces the sensitive term (a first word) “Mr. C. Williams” with the standard token (another word) “<person-3>” because “Mr. C. Williams” is a sensitive term belonging to the category “names of people.” Fahlman, 4:31-37, 5:41-6:16, FIGS. 3-4. That is, the notification information “Mr. C. Williams” matches the first word “Mr. C. Williams” contained in the sensitive terms list and belongs to the same category (type of information) “names of people.” Fahlman, 4:31-37, 5:41-6:16, FIG. 3. This is consistent with the '675 patent, which similarly discloses “type information” corresponding to the category “person” and replacing a first word, e.g., “Mr. Yoshida,” because it likewise corresponds to the category “person.” Ex. 1001, 14:30-41, FIG. 5.

4. Claim 5: “The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice⁴ when the voice spoken by the user is a whisper.”

Raitio discloses the additional limitations of claim 5 because, as explained in [1b], *supra*, the “whispered speech determination module 820” (voice classifying unit) “determine[s] whether the speech input includes a whispered speech input” (a first voice) using “any desired classification methods and/or spectral estimation methods” that “enable[] the detection of speech when the user speaks at an extended distance from [the] digital assistant.” Raitio, [0240], [0242]-[0250], [0253], FIGS. 2A, 8A, 9B, claim 25; Lipoff, ¶120. The “whispered speech determination module 820 can make such determination based on one or more spectrum characteristics, such as the amplitude, the energy, the volume, the slope, or a combination thereof.” Raitio, [0245] (further explaining that the spectrums may be compared to a threshold to classify the input as whispered or non-whispered).

This matches the ’675 patent, which also uses spectrum characteristics to classify a voice as whispered voice, and equates classification of a whispered voice

⁴ Because there is no antecedent basis for “the predetermined voice,” Petitioner interprets “the predetermined voice” as “a predetermined voice.”

with classification of a “predetermined” or “first” voice “V1A.” Ex. 1001, 9:60-10:1 (“In the present embodiment, the voice classifying unit 42 classifies the voice V1 to the first voice V1A when the voice V1 is determined to be a whisper as a result of analysis (determines as the predetermined voice).”), 10:4-8, 10:13-16 (“The voice classifying unit 42 classifies the voice V1 either to the first voice V1A or the second voice V1B by using a peak frequency that is equal to or higher than a predetermined intensity in the spectrum as a feature value.”).

VI. Ground 2: Claims 1-7 Are Obvious over Ocampo and Fahlman

A. Ocampo (Ex. 1005)

Ocampo, filed November 1, 2016, and published May 3, 2018, qualifies as prior art to the ’675 patent under at least 35 U.S.C. § 102(a)(2). Ocampo, cover (22), (43). Ocampo discloses a system and method for generating user command responses based on the content of the command and attributes, such as a whispering tone, of the user’s voice. Ocampo, Abstract, [0003]-[0004], FIGS. 4-5. Figure 4 below depicts an exemplary flowchart of the disclosed method wherein a user device receives a “command to output data” (step 401), data suitable for fulfilling the user command is retrieved (step 404), the data is “converted” into an “audio signal” (step 414), and the audio signal is output to a user through a speaker of the user device (step 416). Ocampo, [0050]-[0053], [0056], [0073], [0077]-[0078], FIG. 4.

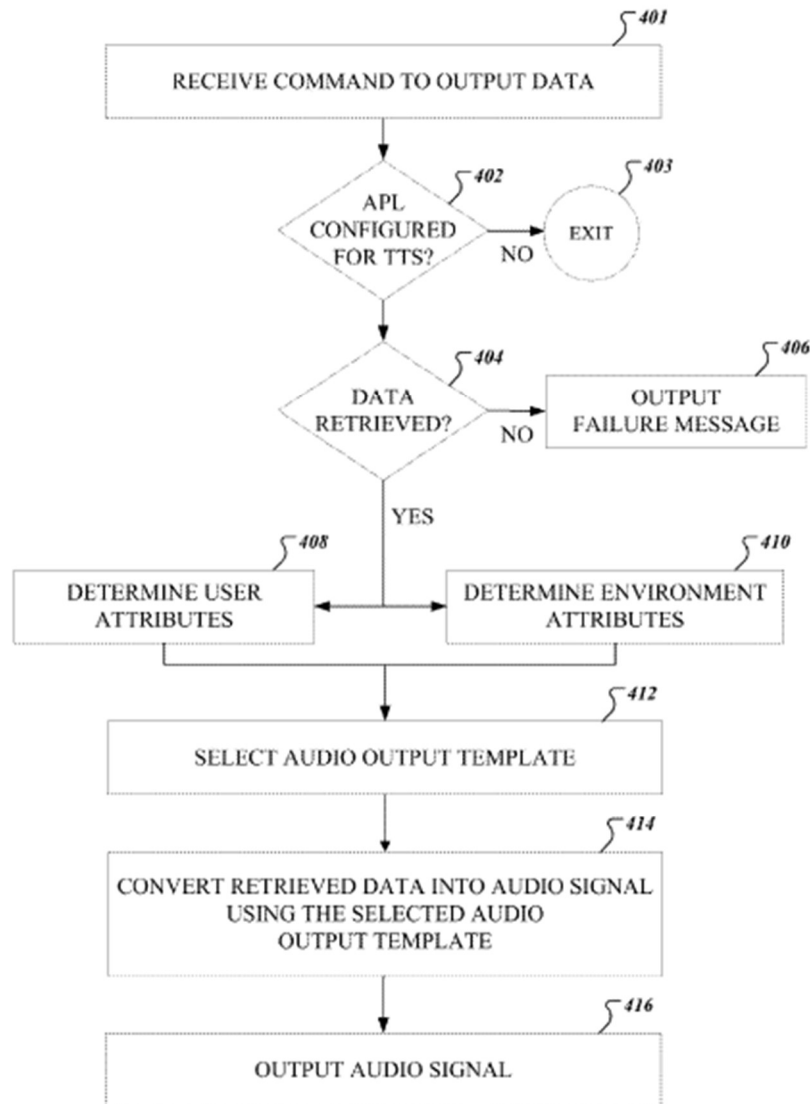


FIG. 4

Ocampo, FIG. 4. Figure 5 below illustrates a user device, such as a “personal digital assistant” or “smart phone,” for implementing the disclosed method.

Ocampo, [0047]-[0048], FIG. 5.

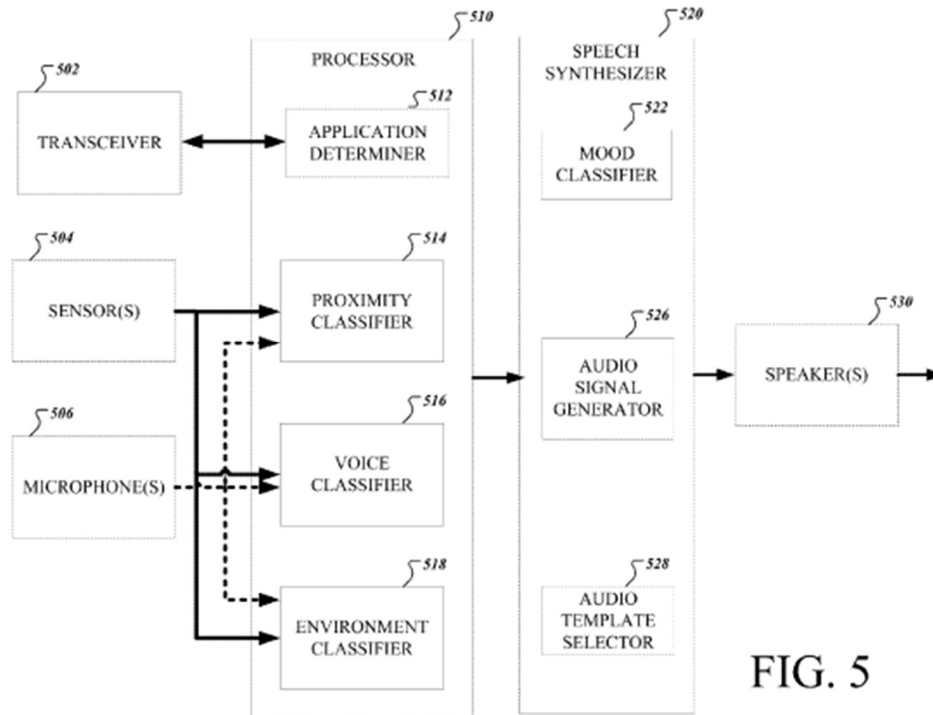


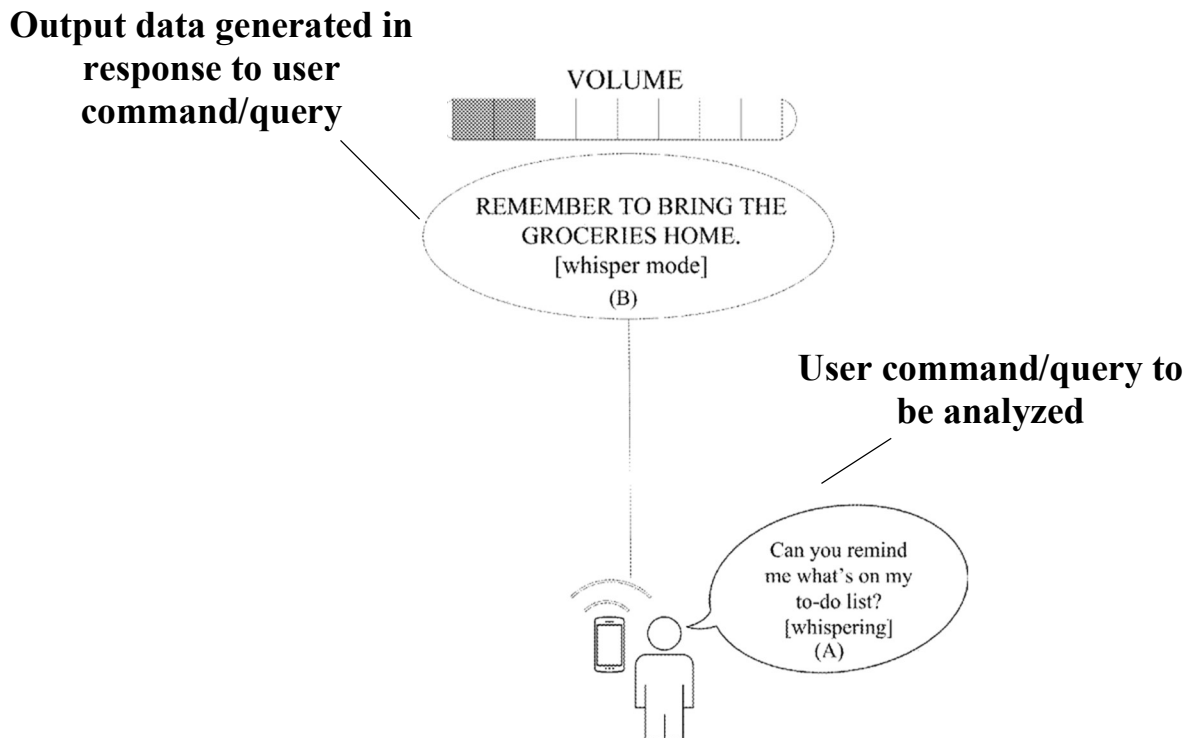
FIG. 5

Ocampo, FIG. 5.

Like the '675 patent, Ocampo's device analyzes the user's command (or query) to determine the user's wishes. Ocampo, [0051], [0053]. Figure 2A illustrates an example where the user asks the device, "Can you remind me what's on my to-do list?" and the device, upon determining the user has submitted a knowledge query/intends to be reminded of their "to-do list," identifies an application to be "used to respond to the user query." Ocampo, [0029]-[0030], [0036]-[0037], [0051], FIG. 2A. The application "search[es] the user's to-do or reminder list to respond to the user query," then an audio signal generator "use[s] any suitable audio synthesizer technique, such as concatenation synthesis, formant

synthesis, articulatory synthesis, and hidden Markov model (HMM)-based synthesis, to convert the retrieved data to an audio signal.” Ocampo, [0036], [0077], FIGS. 2A, 4, 5.

FIG. 2A



Ocampo, FIG. 2A (annotated). The device also classifies the user’s voice input according to its “voice features” so that the outputted audio signal may also reflect, e.g., the user’s tone, loudness, and proximity to the device. Ocampo, [0011], [0023], [0031]-[0032], [0035], [0060]-[0062], [0065], claim 26. In the exemplary scenario of Figure 2A, “the user device determines [based on the voice features] that the user was likely whispering, and selects a voice output template that

corresponds to a whispering audio signal. A voice output template corresponding to a whispering audio signal may include audio signal features such as, for example, a low decibel output, a low volume, and pitch, tone, and frequency corresponding to a whisper.” Ocampo, [0035], FIG. 2A. Thus, when “the data obtained [by the application] in response to the user command is ‘REMEMBER TO BRING THE GROCERIES HOME,’ this data is converted into an audio signal using an audio output template that is selected based on the user attribute indicative of a user having a whispering tone.” Ocampo, [0077], FIGS. 2A, 4 (step 414). The converted output signal is subsequently output through the device’s speakers so that it may be heard by the user. Ocampo, [0078], FIG. 4 (step 416).

As illustrated in Figure 2A, the generated response may also be altered, e.g., to be in a whispering tone depending on whether the device’s “classifiers” “classify [the] audio signals received from a user” as a whisper. Ocampo, [0035], [0037], [0065], FIG. 2A. In this way, the device can account for the user’s need for discrete outputs and privacy. Ocampo, [0039] (recognizing a user might whisper “to avoid potential embarrassment or inconveniencing other people surrounding the user,” and configuring the response by producing “a low volume response” to enhance the “user experience”).

Ocampo is analogous art to the ’675 patent because it is in the same field of endeavor as the ’675 patent: speech processing. *Compare* Ocampo, Abstract, with

Ex. 1001, Abstract; Lipoff, ¶66. Furthermore, Ocampo is directed to the same problem as the '675 patent: analyzing voice input commands to generate responses. *Compare* Ocampo, Abstract, *with* Ex. 1001, Abstract; Lipoff, ¶66.

B. Independent Claims

1. Claim 1

Ocampo discloses or suggests every element of claim 1. While Ocampo does not expressly disclose performing word replacement, as recited in element [1f], *infra*, Fahlman does.

a. [1preamble]: “An output-content control device comprising:”

Ocampo discloses [1preamble] because it discloses a “user device” for “convert[ing] the content in [a] received message to an audio signal and control[ling] the output of [an] audio signal to be at a volume proportional to [a] determined proximity indicator.” Ocampo, [0024]; Lipoff, ¶137. Figure 2A below illustrates a user speaking the query, “Can you remind me what’s on my to-do list?” and a user device generating responsive output content, “REMEMBER TO BRING THE GROCERIES HOME.” Ocampo, [0029]-[0037], [0077], FIG. 2A.

FIG. 2A



Ocampo, FIG. 2A.

- b. [1a]: “a voice acquiring unit configured to acquire a voice spoken by a user;”

Ocampo discloses [1a] because it discloses a processor for “obtaining audio signal data from a first **microphone**” and “obtaining audio signal data from a

second **microphone**.” Ocampo, [0011], FIG. 5; *see also* Ocampo, [0031] (“From the actuated microphones, the user device may obtain samples of the user’s voice.”), [0092] (“The processes and logic flows described in this specification may be performed by one or more programmable processors executing one or more computer programs to perform actions by operating on input data and generating output.”); Lipoff, ¶138.

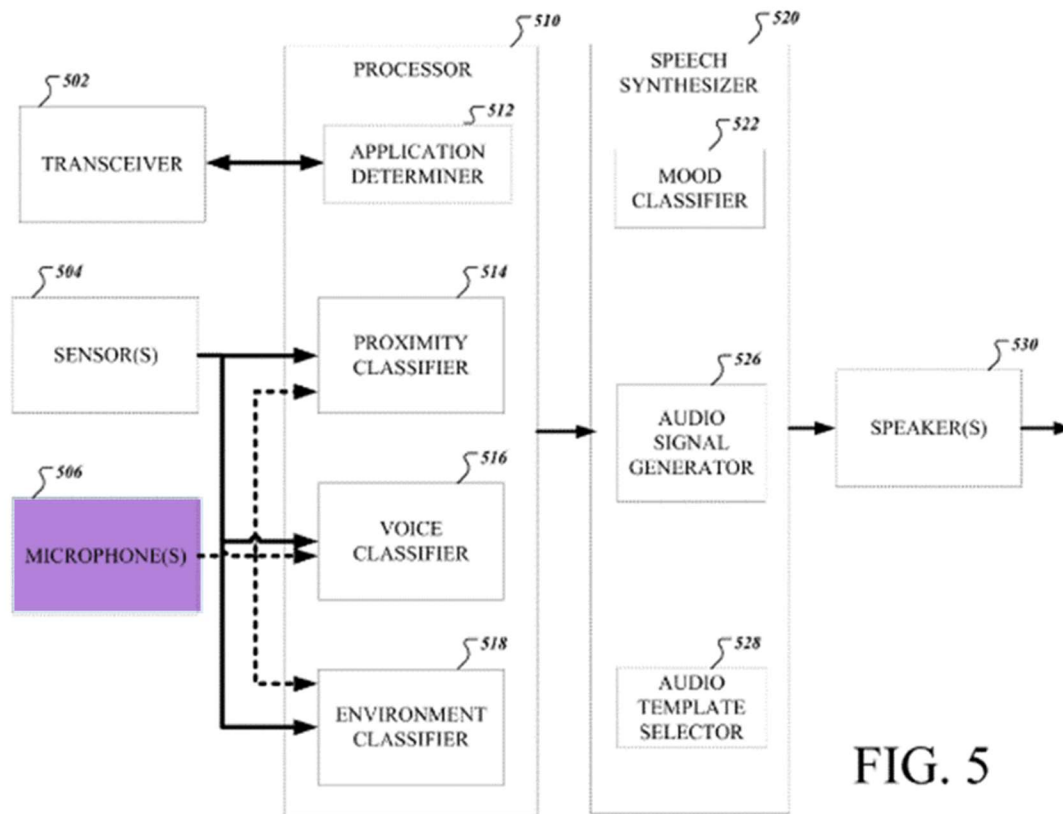


FIG. 5

Ocampo, FIG. 5 (annotated).

This matches the '675 patent, which similarly explains that its voice acquiring unit is a CPU “unit” that acquires a voice “detected by [a] voice detecting unit” such as a microphone. Ex. 1001, 3:27-28, 3:55- 4:2.

- c. **[1b]: “a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”**

Ocampo discloses or renders obvious [1b] because it discloses “**proximity**,” “**voice**,” and “**mood**” classifiers (**collectively a voice classifying unit**) for classifying a user’s voice as a “whisper[]” (first voice) or non-whisper (second voice) based on a determined “likely location and a likely distance of the user” (calculated distance between the user and output-content control device). Ocampo, [0011], [0047], [0062], [0065], claim 26; Lipoff, ¶140.

As illustrated in annotated Figure 5 below, Ocampo’s user device includes various “classifiers” (collectively a voice classifying unit) configured to “classify” user “voice samples” based on “one or more voice features” such as “a pitch, tone, frequency, and amplitude of an audio signal corresponding to the user’s voice.” Ocampo, [0031]-[0032], [0065] (“By using classifiers to classify audio signals received from a user and determine user attributes, *the likely user temperament, such as whether a user is whispering*, shouting, happy, sad, or excited, may be determined.”).

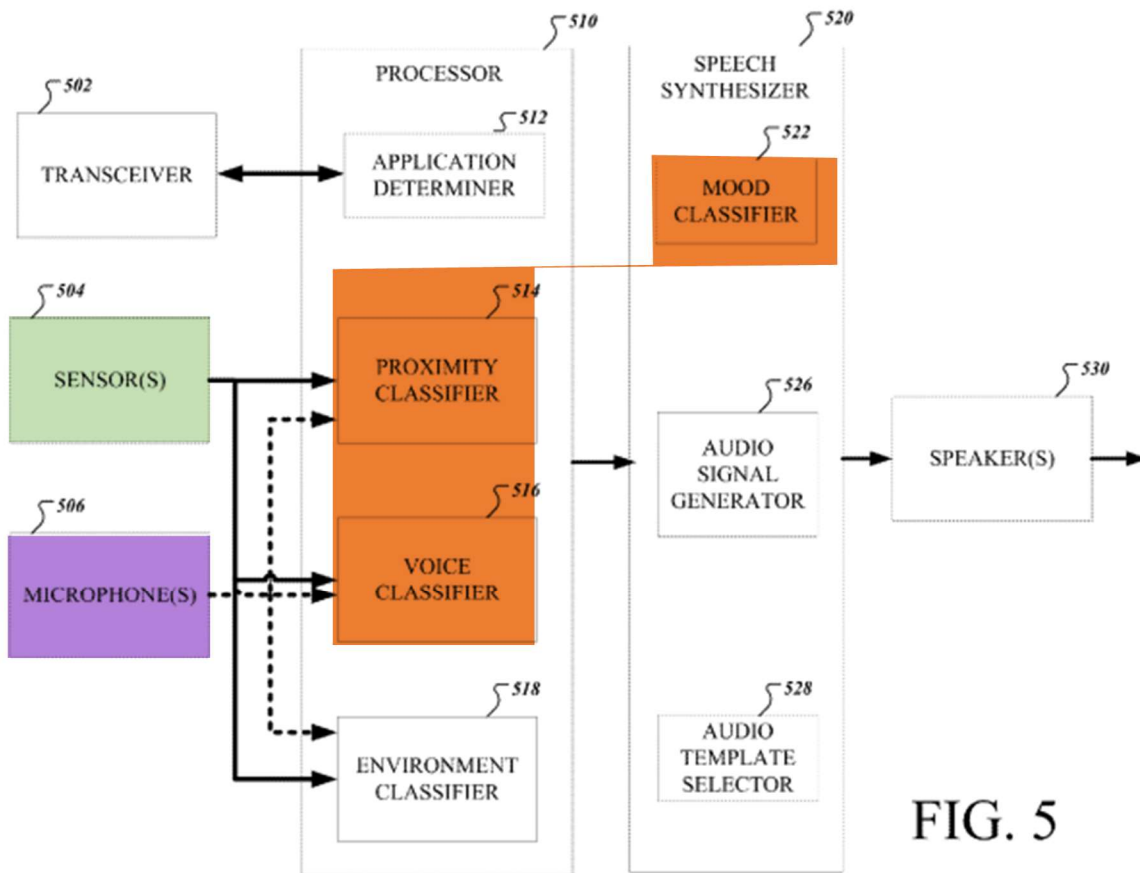


FIG. 5

Ocampo, FIG. 5 (annotated).

For example, the “**voice classifier**” receives “extracted amplitude and frequency data to determine a *pitch and tone* of the user’s voice” from the user device’s **microphones** and **sensors**; a “**mood classifier**” “predict[s] the likely mood of the user *based on the* [determined] *pitch, tone*, amplitude, and frequency data of the audio signal”; and a “**proximity classifier**” determines the user’s “location” and “distance” with respect to the user device using input from the device’s microphones and sensors. Ocampo, [0011], [0023], [0060]-[0062], [0065]; *see also*

Ocampo, [0087] (explaining that the voice classifier may be “trained with training data for identifying voice features such as pitch and tone,” and the mood classifier may be “trained with training data for identifying user moods”), [0088] (explaining that the proximity classifier “may be trained to interpret audio signal data and patterns from one or more microphones and data from sensors to determine the likely location and position of a user relative to the user device”).

Ocampo further discloses that determining “the tone of voice of the user associated with the user device or the voice volume of the user associated with the user device” includes “*determining a likely location and a likely distance of the user based on the sensor data, audio signal data from the first microphone, and the audio signal data from the second microphone.*” Ocampo, claim 26, [0011], [0023]; *see also* Ocampo, [0060]-[0062] (explaining that the proximity classifier calculates the user’s “likely” location and distance). A first voice is classified as a “whisper” when the “voice features” correspond to, e.g., “a low decibel output, a low volume, and pitch, tone, and frequency.” Ocampo, [0035].

Thus, a POSITA would have understood that Ocampo’s proximity, voice, and mood classifiers work as a unit (the **proximity classifier** calculating user location distance, the **voice classifier** using the calculated distance to determine user tone and pitch, and the **mood classifier** using the determined tone and pitch to

determine whether a user is whispering) to classify the user's voice as a whisper (first voice) or non-whisper (second voice). Lipoff, ¶¶140-144.

- d. **[1c]: “an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;”**

Ocampo discloses [1c] because the user device includes an “[application determiner](#)” (intention analyzing unit) configured to “access[] ... mapping” information correlating the user command to a “particular application” and “determine which application to use to process or respond to the command.” Ocampo, [0051]-[0053]; Lipoff, ¶145.

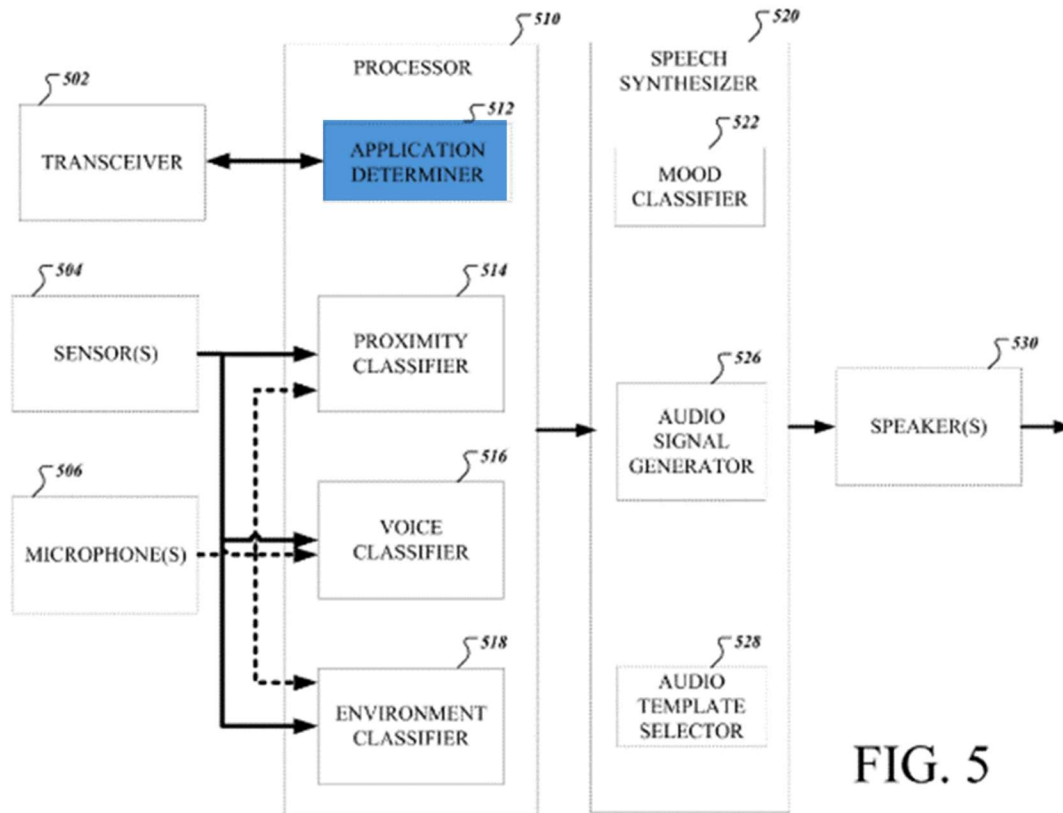
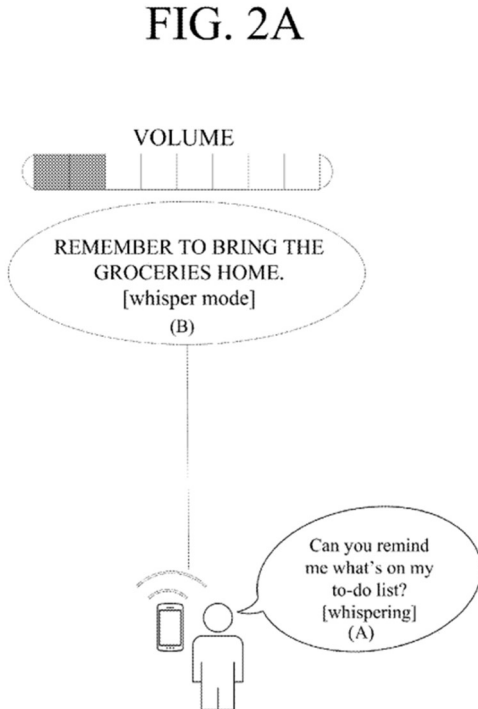


FIG. 5

Ocampo, FIG. 5 (annotated).

As Ocampo explains, user “commands may be classified and mapped to a particular application.” Ocampo, [0051]. The applications “process or respond to the [user] command[/query],” and the application determiner “*accesses the mapping information* to determine which application to use to process or respond to the command.” Ocampo, [0051]. For example, “if a command to output an electronic or text message is received,” the application determiner accesses the mapping information to determine that the command is “a text messaging output command” and selects “a messaging application ... to output the received

message”—that is, the application determiner analyzes the command (e.g., its associated classification/mapping) and detects that the user wishes to receive a text message output. Ocampo, [0051], [0053]; Lipoff, ¶146. If a “command corresponding to a user query [is] *classified as a knowledge query*,” the application determiner accesses the classification information to select, e.g., a “browser application ... to respond to the query with data retrieved from a network, such as the Internet.” Ocampo, [0051]. Thus, the application determiner analyzes the query (e.g., its associated classification/mapping) and detects that the user wishes to receive particular knowledge. Ocampo, [0051], [0053]; Lipoff, ¶¶146-147. Figure 2A illustrates an example where the voice input is the knowledge query, “Can you remind me what’s on my to-do list?” and the application determiner, upon determining the user has submitted a knowledge query/intends to be reminded of their “to-do list,” identifies the application to be “used to respond to the user query.” Ocampo, [0029]-[0030], [0036]-[0037], [0051], FIG. 2A.



Ocampo, FIG. 2A.

Accordingly, a POSITA would have understood that the application determiner is an intention analyzing unit because it analyzes the user's query to determine that the user intends to know what is on their "to-do list" (the wish to be acquired by the user) so that the user device can retrieve the information necessary to respond to the wish, as further explained in [1d], *infra*. Ocampo, [0036]-[0037], [0051], FIG. 2A; Lipoff, ¶¶145-148.

This example in Ocampo is consistent with the '675 patent, which explains that its intention analyzing unit recognizes that the information the user wishes to

acquire is about a user's schedule, when the user's query corresponds to the text, "today's schedule is." Ex. 1001, 4:15-5:3.

- e. **[1d]: "a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and"**

Ocampo discloses [1d] because an application (notification-information acquiring unit) selected by the application determiner is configured to "retrieve data for processing or responding to the [user] command" (notification information). Ocampo, [0056], [0036]; Lipoff, ¶150.

As explained in Ground 2, [1c], *supra*, the user device includes an "application determiner" (intention analyzing unit) configured to "determine which application to use to process or respond to the command." Ocampo, [0051]-[0053]; Lipoff, ¶151. The selected "application may be preconfigured to *retrieve data* from an application database or application server," or "*retrieve data* from various suitable data sources in response to the same [user] command." Ocampo, [0056]. The data corresponds to notification information, and the exact source of the data "depends on various factors[,] including the type of application and type of command." Ocampo, [0056]. For example, the data itself "may be retrieved in various suitable ways[,] including ... communicating with a network, such as the Internet, to retrieve data, or communicating with a server, database, or storage

device to retrieve data.” Ocampo, [0056], FIG. 4. This data retrieval operation is illustrated in annotated Figure 4 below.

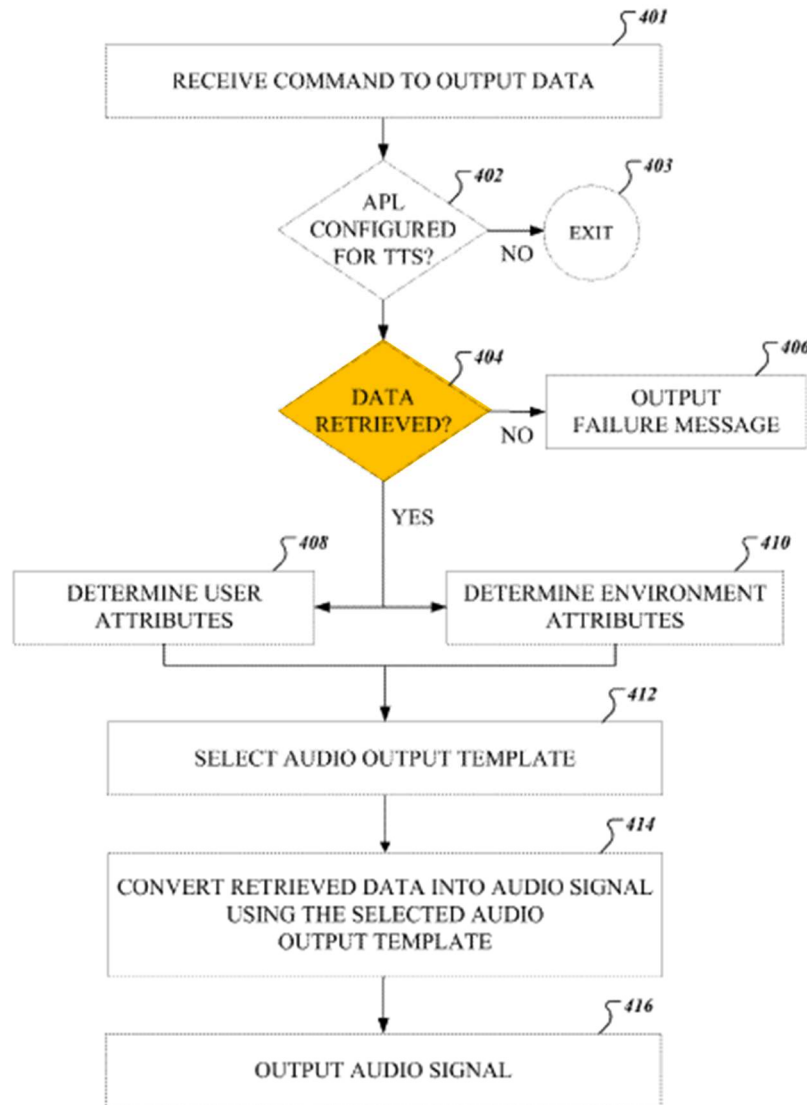


FIG. 4

Ocampo, FIG. 4 (annotated).

In the embodiment illustrated in Figure 2A, the application “search[es] the user’s to-do or reminder list to respond to the user query. This information may be

obtained by communicating with a server in a network or retrieving data stored in a storage device. The storage device may be integrated into the user device or attached to the user device.” Ocampo, [0036], FIG. 2A.

Thus, a POSITA would have understood that the application is a “notification-information acquiring unit” because it retrieves data from an application database, application server, or various suitable data sources in response to the user’s command and uses the retrieved data to configure the user response, as further explained in [1e], *infra*. Ocampo, [0056]; Lipoff, ¶¶150-152.

- f. **[1e]: “an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information,”**

Ocampo discloses [1e] because its user device includes an “audio signal generator” (output-content generating unit) for “convert[ing]” responsive the information acquired by the application “into an audio signal.” Ocampo, [0037], [0077]; Lipoff, ¶153.

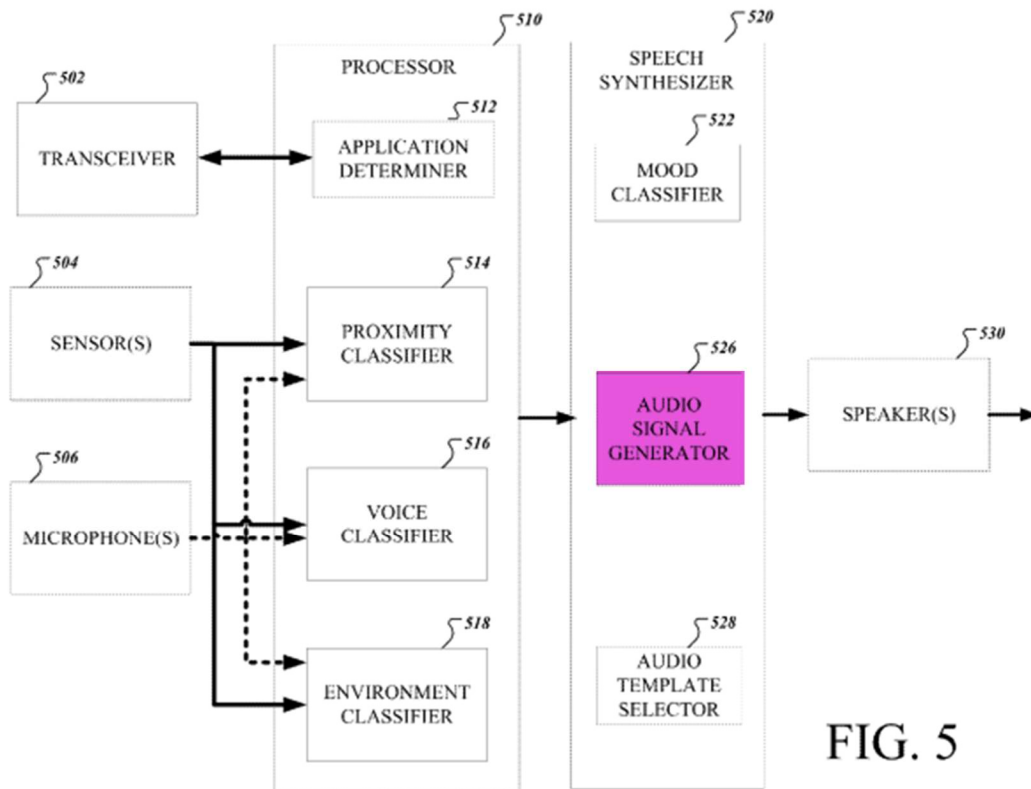


FIG. 5

Ocampo, FIG. 5 (annotated).

As explained for Ground 2, [1d], *supra*, Ocampo's application (notification-information acquiring unit) retrieves data from an application database, application server, or other suitable data sources in response to the user's command, so that the retrieved data may be used to configure a user response. Ocampo, [0056], FIG. 4; Lipoff, ¶154.

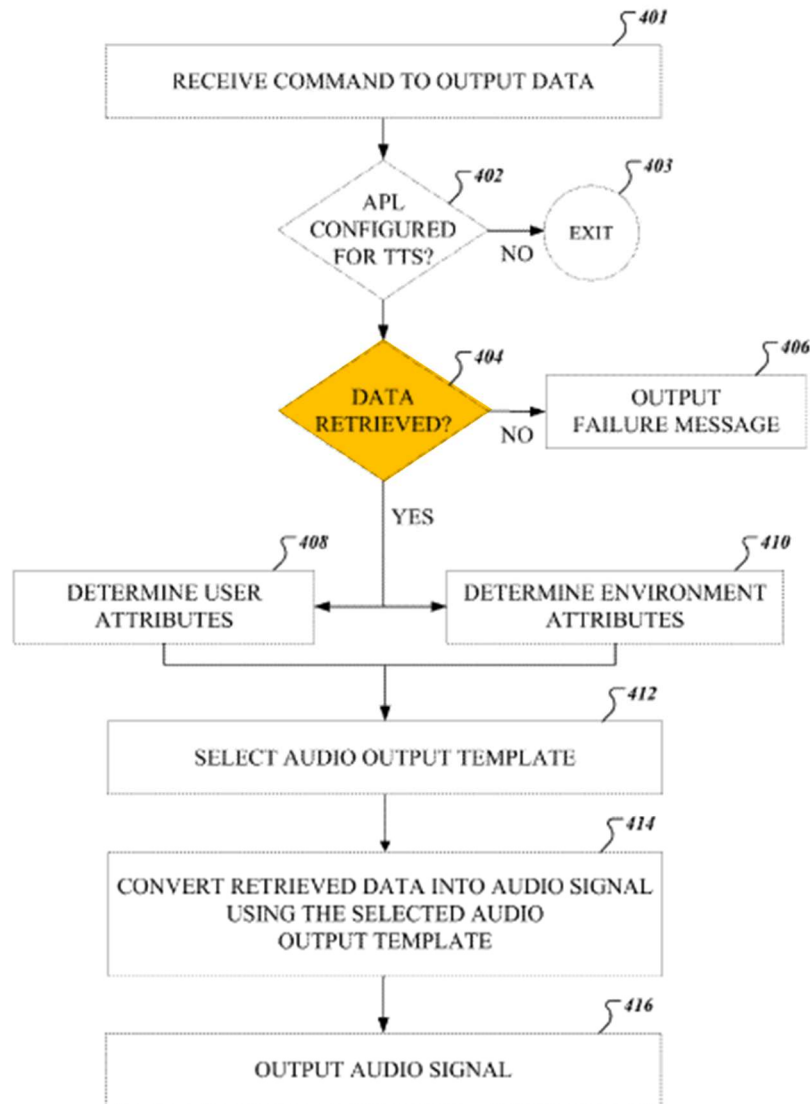


FIG. 4

Ocampo, FIG. 4 (annotated). Ocampo further explains that the retrieved data is sent to an “audio signal generator” (output-content generating unit) where it is converted into an audio signal using any of a variety of techniques. Ocampo, [0077], FIGS. 4-5. For example, the audio signal generator “may use any suitable audio synthesizer technique, such as concatenation synthesis, formant synthesis,

articulatory synthesis, and hidden Markov model (HMM)-based synthesis, to *convert the retrieved data to an audio signal.*” Ocampo, [0077], FIG. 4.

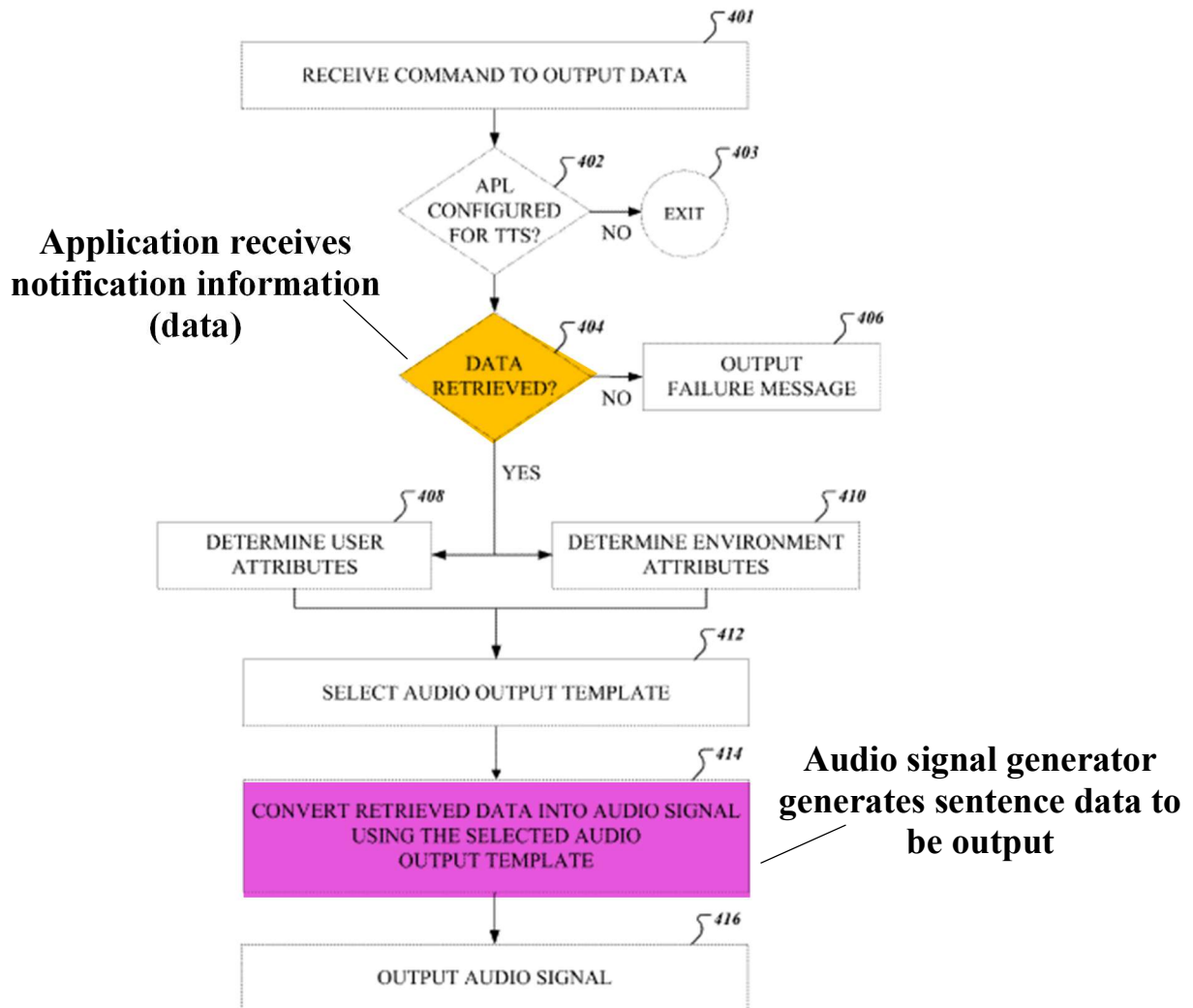
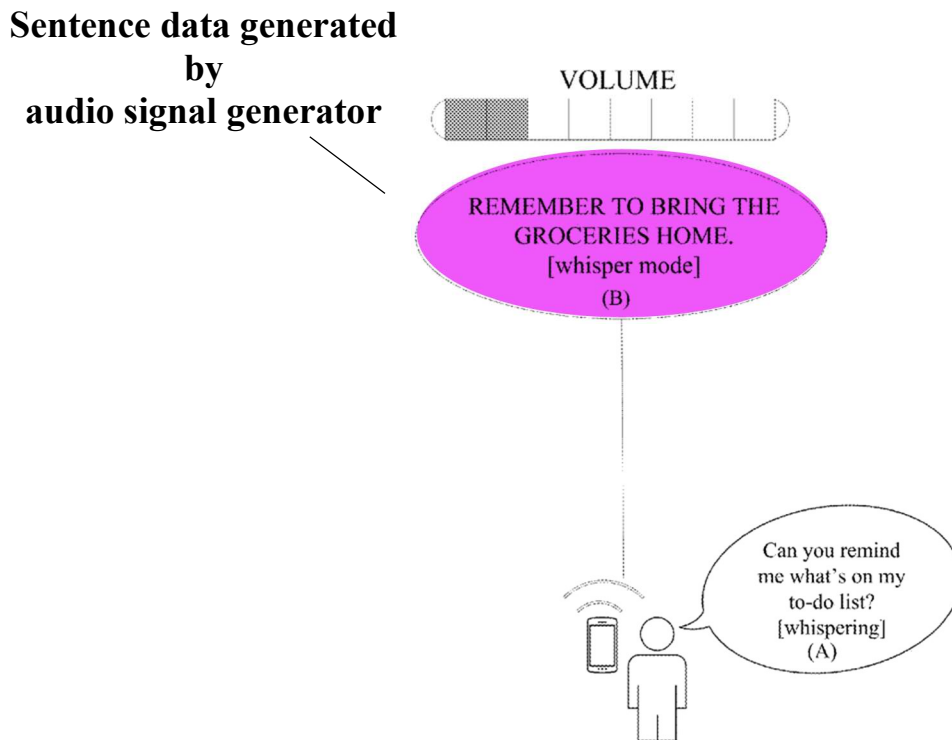


FIG. 4

Ocampo, FIG. 4 (annotated). Referring back to the exemplary scenario of Figure 2A, “if the data obtained [by the application] in response to the user command is ‘REMEMBER TO BRING THE GROCERIES HOME,’ this data is

converted into an audio signal using an audio output template that is selected based on the user attribute indicative of a user having a whispering tone.” Ocampo, [0077], FIG. 2A; *see also* Ocampo, [0047], [0050], [0056], FIG. 4 (describing “a method for providing a dynamic TTS [(text-to-speech)] output”). The converted output signal is subsequently output through the device’s speakers so that it may be heard by the user. Ocampo, [0078], FIG. 4 (steps 414-416).

FIG. 2A



Ocampo, FIG. 2A (annotated).

Thus, Ocampo discloses an output-content generating unit (the “audio signal generator”) configured to generate an output sentence (e.g., “REMEMBER TO

BRING THE GROCERIES HOME”) as sentence data to be output to the user based on the notification information. Ocampo, [0077], FIGS. 2A, 4-5. Fahlman similarly discloses a system and method for generating an output message (sentence), as explained in further detail in [1f], *infra*. Fahlman, 3:35-60.

- g. **[1f]: “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word, and to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”**

Ocampo-Fahlman renders obvious [1f]. Lipoff, ¶¶157-165.

Ocampo at least suggests [1f] because Ocampo teaches modifying the generated response (content information of the notification information) when the voice corresponds to a “whisper” (first voice). Ocampo, [0035], [0037]; Lipoff, ¶158.

Ocampo further discloses that responding in a manner commensurate with the user’s attributes (e.g., “tone” and “mood”) “offers several advantages.” Ocampo, [0038]-[0039]. For example, a user might whisper “to avoid potential embarrassment or inconveniencing other people surrounding the user”—selecting

an audio template that appreciates this, e.g., by producing “a low volume response,” enhances the “user experience.” Ocampo, [0039].

Ocampo doesn’t expressly disclose performing word replacement based on a determination that the user whispered/may be embarrassed by a loud response (low amplitude, volume, etc.), but Fahlman does.

Fahlman teaches “a method and apparatus for transforming a message containing sensitive information” by replacing one or more “terms containing sensitive information” with a “standard token” to generate a “sanitized message.” Fahlman, Abstract, 1:7-12, 3:24-26, 3:35-53. “For example, the sensitive term ‘Mr. Johnson’ is replaced by the standard token <person-1>, and the term ‘Jul. 1, 1997’ is replaced by <date-1>.” Fahlman, 3:48-52. As Fahlman explains, “[t]he sanitized message” which may be text or voice, “is then transmitted with a low level of security since the sanitized message only contain[s] standard tokens, not sensitive information.” Fahlman, 3:29-33, 3:54-60.

i. Motivation to Combine Ocampo and Fahlman

A POSITA would have appreciated that Fahlman’s word replacement would have been useful for generating responses that obviate the embarrassment concerns of Ocampo’s user. Lipoff, ¶160. For example, a POSITA would have appreciated that in addition to whispering a generated response (as described above), sanitizing the response to remove any embarrassing (sensitive) information would have

further enhanced the user's experience by reducing the likelihood that others in the user's vicinity would also understand the response.

Thus, a POSITA would have been motivated to implement Fahlman's word replacement algorithms in Ocampo's template selector (output-content generating unit) to remove any embarrassing or sensitive information from the response that the user may not want others in his/her proximity to hear. Lipoff, ¶¶160-161.

ii. Expectation of Success

A POSITA would have had a reasonable expectation of success in implementing Fahlman's algorithm because it would have required only routine skill to implement. Lipoff, ¶¶162-165.

As Ocampo explains, "all of the functional operations and/or actions described in [Ocampo] may be implemented in digital electronic circuitry, or in computer software, firmware, or hardware," and "[t]he processes and logic flows described in [Ocampo] may be performed by one or more programmable processors executing one or more computer programs to perform actions by operating on input data and generating output." Ocampo, [0090], [0092]. Indeed, Ocampo's "audio template selector" is a program in the user device's speech synthesizer processor. Ocampo, [0047].

Fahlman similarly discloses that "a suitable computer program algorithm" may be used to generate the sanitized message. Fahlman, 3:35-53. For example, an

algorithm may be used to identify the sensitive terms contained in the message using, e.g., a “defined list of sensitive terms, such as known places, known names, dollar (\$) signs, etc.” Fahlman, 3:35-47, 6:29-34. And “another subroutine or various combination[s] of algorithms” may be used “for replacing the identified sensitive terms with standard tokens, thereby creating a sanitized message.” Fahlman, 6:35-38.

Thus, a POSITA would have had a reasonable expectation of success in implementing Fahlman’s word replacement algorithms in Ocampo’s template selector because it would have merely required the integration of computer algorithms and routine skill to do so. Lipoff, ¶¶162-165.

2. Claim 6

a. [6preamble]: “An output-content control method comprising:”

Ocampo discloses [6preamble] because it discloses “a method for providing a dynamic TTS [(text-to-speech)] output.” Ocampo, Abstract, [0047], FIG. 4.

Figure 4 below illustrates a method implemented by the system of Figure 5 (*see* Ground 2, [1preamble], *supra*), whereby an “OUTPUT AUDIO SIGNAL” is generated (Ocampo, FIG. 4 (step 416)) responsive to “RECEIV[ING] [A] COMMAND TO OUTPUT DATA” (Ocampo, FIG. 4 (step 401)). Ocampo, [0047], [0050], [0077]-[0078].

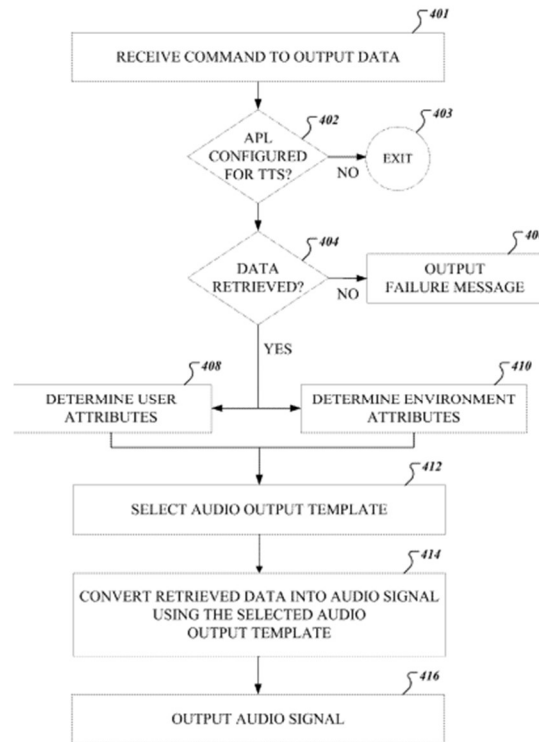


FIG. 4

Ocampo, FIG. 4.

b. [6a]: “acquiring a voice spoken by a user;”

Ocampo discloses [6a] for reasons discussed for Ground 2, [1a], *supra*.

Lipoff, ¶184.

c. [6b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”

Ocampo discloses [6b] for reasons discussed for Ground 2, [1b], *supra*.

Lipoff, ¶185.

- d. **[6c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”**

Ocampo discloses [6c] for reasons discussed for Ground 2, [1c], *supra*.

Lipoff, ¶186.

- e. **[6d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”**

Ocampo discloses [6d] for reasons discussed for Ground 2, [1d], *supra*.

Lipoff, ¶187.

- f. **[6e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and”**

Ocampo-Fahlman renders obvious [6e] for reasons discussed for Ground 2, [1e]-[1f], *supra*. Lipoff, ¶188.

- g. **[6f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”**

Ocampo-Fahlman renders obvious [6f] for reasons discussed for Ground 2, [1e]-[1f], *supra*. Lipoff, ¶189.

3. Claim 7

- a. [7preamble]: “A non-transitory storage medium that stores an output-content control program that causes a computer to execute:”**

Ocampo discloses [7preamble] because its disclosed processes “may be implemented as one or more computer program products, for example, one or more modules of computer program instructions encoded on a computer readable medium for execution by, or to control the operation of, [a] data processing apparatus,” and “[c]omputer-readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory devices, including by way of example semiconductor memory devices, for example, EPROM, EEPROM, and flash memory devices; magnetic disks, for example, internal hard disks or removable disks; magneto optical disks; and CD ROM and DVD-ROM disks. The processor and the memory may be supplemented by, or incorporated in, special purpose logic circuitry.” Ocampo, [0090]-[0094], claim 27.

- b. [7a]: “acquiring a voice spoken by a user;”**

Ocampo discloses [7a] for reasons discussed for Ground 2, [1a], *supra*. Lipoff, ¶191.

- c. **[7b]: “calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;”**

Ocampo discloses [7b] for reasons discussed for Ground 2, [1b], *supra*.

Lipoff, ¶192.

- d. **[7c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”**

Ocampo discloses [7c] for reasons discussed for Ground 2, [1c], *supra*.

Lipoff, ¶193.

- e. **[7d]: “acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and”**

Ocampo discloses [7d] for reasons discussed for Ground 2, [1d], *supra*.

Lipoff, ¶194.

- f. **[7e]: “generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and”**

Ocampo-Fahlman renders obvious [7e] for reasons discussed for Ground 2,

[1e]-[1f], *supra*. Lipoff, ¶195.

- g. [7f]: “generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.”**

Ocampo-Fahlman renders obvious [7f] for reasons discussed for Ground 2, [1e]-[1f], *supra*. Lipoff, ¶196.

C. Dependent Claims

- 1. Claim 2: “The output-content control device according to claim 1, wherein the notification information includes type information indicating a type of the content information, and the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be [re]placed with the other word based on a word included in the content information and the type information that is associated with the content information.”**

Ocampo-Fahlman renders obvious claim 2 as explained below. Lipoff, ¶¶166-173.

“the notification information includes type information indicating a type of the content information”

As explained for Ground 2, [1d], *supra*, Ocampo discloses acquiring notification information to be notified to the user, and that the notification information includes content information, i.e., the words to be conveyed to the user based on the user’s query. Ocampo, [0036], [0056], FIG. 2A. The content information in the output also corresponds to type information. For example, the word/content “HOME” in the notification information “REMEMBER TO BRING

THE GROCERIES HOME” corresponds to the type of information *place*.

Ocampo, [0077]. This matches the ’675 patent, which explains that “type information” describes a category of information such as “person,” “date,” or place, while “content information” describes specific content in the category, e.g., “Mr. Yamada,” “Mar. 20, 2020,” or “Tokyo station.” Ex. 1001, 5:31-53, 8:21-44, FIGS. 4-5.

Fahlman likewise discloses that its notification information includes not only the words to be notified to the user (content information) but categories (type information) indicating a type of the content. For example, the content information “Mr. C. Williams” corresponds to the category (type information) “names of people or organizations,” while the content information “Jul. 7, 1997” corresponds to the category “dates & other time description.” Fahlman, 4:31-37, 5:41-6:16, FIG. 3.

Categorizing the content information by type enables the device to identify which words to output and which words to replace in the sanitized message. Fahlman, 4:20-46; *see also* Raitio, [0232] (describing a “structured query” that separates the words in user queries and responses into distinct categories to facilitate query fulfillment).

Thus, to the extent Ocampo does not expressly disclose that its notification information includes type information, a POSITA would have been motivated to

modify the notification information such that it includes type information indicating a type of the content information, to facilitate user query fulfillment by allowing the device to easily identify which words to output in response to the user's query. Fahlman, 4:20-46; Raitio, [0232]. A POSITA would have had a reasonable expectation of success in making such a modification for reasons articulated in Ground 2, [1f], *supra*.

“the output-content generating unit further configured to select, on generating the first output sentence, at least one word to be [re]placed with the other word based on a word included in the content information and the type information that is associated with the content information”

As explained for Ground 2, [1f], *supra*, Fahlman teaches and renders obvious replacing one or more “terms containing sensitive information” with a “standard token” to generate a “sanitized message.” Fahlman, Abstract, 1:7-12, 3:24-26, 3:35-53. Fahlman further teaches that the sensitive term (selected word to be replaced) is based on a word included in the content information and associated type information.

For example, when the notification information is “I have interviewed Mr. C. Williams and believe that we should hire him at an annual salary of \$80,000 commencing on Jul. 7, 1997,” the subroutine replaces the sensitive term (at least one word) “Mr. C. Williams” with the standard token (another word) “<person-3>”

based on its classification as a sensitive term (type information) and its inclusion in the content information. Fahlman, 5:41-6:16, FIGS. 3-4.

Thus, Ocampo's output-content generating unit, as modified by Fahlman, teaches the output-content generating unit further configured to select, on generating the first output sentence, at least one word to be replaced with the other word based on a word included in the content information and the type information that is associated with the content information.

2. **Claim 3: "The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word."**

Ocampo-Fahlman renders claim 3 obvious. Lipoff, ¶¶174-176.

As explained in Ground 1, claim 3, *supra*, Fahlman further explains that its computer program/subroutine (output-content generating unit) is configured to read "a defined list of sensitive terms" and corresponding "standard tokens" (relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word) and to replace the sensitive term (the first word matching a word in the notification

information) with the standard token (the second word). Fahlman, 3:34-53, FIGS. 3-4.

For example, when the notification information is “I have interviewed Mr. C. Williams and believe that we should hire him at an annual salary of \$80,000 commencing on Jul. 7, 1997,” the subroutine identifies the sensitive term “Mr. C. Williams” from its “defined list of sensitive terms” and replaces the sensitive term (a word included in the notification information/the first word) “Mr. C. Williams” with the standard token (second word) “<person-3>.” Fahlman, 4:44-47, 5:41-6:16, FIGS. 3-4.

3. **Claim 4: “The output-content control device according to claim 3, wherein the relationship information further includes information of type information indicating a type of the first word, and the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word.”**

Ocampo-Fahlman renders claim 4 obvious because, as explained in Ground 1, claim 4, *supra*, Fahlman further discloses that the relationship information includes category information indicating a type of the sensitive term (type information indicating a type of the first word). Lipoff, ¶¶177-179.

For example, the relationship information includes the categories “names of people or organizations.” Fahlman, 4:31-37, FIG. 3. Fahlman’s computer program/subroutine (output-content generating unit) is also configured to select a sensitive term for replacement with a standard token first (a word included in the notification information as the word to be replaced with another word) when the sensitive term belongs to the same category (type) as the term in the notification information, and the word in the notification information matches the sensitive term.

For example, when the notification information is “I have interviewed Mr. C. Williams and believe that we should hire him at an annual salary of \$80,000 commencing on Jul. 7, 1997,” the subroutine replaces the sensitive term (a first word) “Mr. C. Williams” with the standard token (another word) “<person-3>” because “Mr. C. Williams” is a sensitive term belonging to the category “names of people.” Fahlman, 4:31-37, 5:41-6:16, FIGS. 3-4. That is, the notification information “Mr. C. Williams” matches the first word “Mr. C. Williams” contained in the sensitive terms list and belongs to the same category (type of information) “names of people.” Fahlman, 4:31-37, 5:41-6:16, FIG. 3 This is consistent with the ’675 patent, which similarly discloses “type information” corresponding to the category “person” and replacing a first word, e.g., “Mr. Yoshida,” because it likewise corresponds to the category “person.” Ex. 1001, 14:30-41, FIG. 5.

4. Claim 5: “The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.”

Ocampo discloses the additional limitations of claim 5 because, as explained in Ground 2, [1b], *supra*, its classifiers “classify audio signals received from a user and determine ... *whether a user is whispering.*” Ocampo, [0035], [0037], [0065], FIG. 2A; Lipoff, ¶180. This is consistent with the ’675 patent, which also equates classification of a whispered voice with classification of a “predetermined” or “first” voice “V1A.” Ex. 1001, 9:60-10:1 (“In the present embodiment, the voice classifying unit 42 classifies the voice V1 to the first voice V1A when the voice V1 is determined to be a whisper as a result of analysis (determines as the predetermined voice).”), 10:4-8.

Ocampo also discloses the additional limitations of claim 5 because it discloses that “[d]ata indicative of the voice features and classification may be added as user attributes to a user profile stored in a user database” such that, “[w]hen a subsequent audio instruction is received by the user device, the user device may determine whether the received audio instruction has been issued by the user by comparing voice features of the audio instruction with stored voice features associated with the user.” Ocampo, [0032], [0041], [0067].

As Ocampo explains, the user device may “receive[] samples of the user’s voice, process[] the voice samples, and store[] data for voice recognition” and “security” purposes. Ocampo, [0041]. For example, the vehicle control module device “may process a user’s voice sample to detect pitch, tone, frequency, and enunciations of the user and store these voice features as user attributes in a user profile. When a subsequent audio instruction is received by the user device, the user device may determine whether the received audio instruction has been issued by the user by comparing voice features of the audio instruction with stored voice features associated with the user.” Ocampo, [0041]. Thus, Ocampo also discloses that its classifier is configured to determine that the voice belongs to a particular user (a predetermined voice).

VII. Mandatory Notices

A. Real Parties-in-Interest Under 37 C.F.R. § 42.8(b)(1)

Google LLC is the real party-in-interest for this Petition.⁵

B. Related Matters Under 37 C.F.R. § 42.8(b)(2)

To the best of Petitioner’s knowledge, the ’675 patent is or has been involved in the following cases:

⁵ Google LLC is a subsidiary of XXVI Holdings Inc.; a subsidiary of Alphabet Inc.

XXVI Holdings Inc. and Alphabet Inc. are not real parties-in-interest here.

Name	Number	Court	Filed
<i>SoundClear Technologies LLC v. Google, LLC</i>	1:24-cv-01281	E.D. Va.	July 25, 2024
<i>SoundClear Technologies LLC v. Amazon.com, Inc. et al.</i>	1:24-cv-01283	E.D. Va.	July 25, 2024
<i>SoundClear Technologies LLC v. Google, LLC</i>	3:24-cv-00540	E.D. Va.	July 25, 2024
<i>SoundClear Technologies LLC v. Amazon.com, Inc. et al.</i>	25-1288	CAFC	Dec. 16, 2024

C. Lead and Backup Counsel Under 37 C.F.R. § 42.8(b)(3)

Petitioner identifies the following lead and backup counsel:

Lead Counsel	Backup Counsel
Erika H. Arner (Reg. No. 57,540) erika.arner@finnegan.com Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 1875 Explorer Street Suite 800 Reston, VA 20190-6023 Tel.: 571-203-2754 Fax: 202-408-4400	Cory C. Bell (Reg. No. 75,096) cory.bell@finnegan.com Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 2 Seaport Lane, 6th Floor Boston, MA 02210-2001 Tel.: 617-646-1641 Fax: 202-408-4400

Lead Counsel	Backup Counsel
	Dan C. Tucker (Reg. No. 62,781) daniel.tucker@finnegan.com Michael V. Young (Reg. No. 61,180) michael.young@finnegan.com Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 1875 Explorer Street Suite 800 Reston, VA 20190-6023 Tel.: 571-203-2700 Fax: 202-408-4400 Safiya Aguilar (Reg. No. 77,212) safiya.aguilar@finnegan.com Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 901 New York Avenue, NW Washington, DC 20001-4413 Tel.: 202-408-4160 Fax: 202-408-4400

D. Service Information Under 37 C.F.R. § 42.8(b)(4)

Please address correspondence to lead and backup counsel at the addresses above and Google-Soundclear-IPRs@finnegan.com. Petitioner consents to electronic service by e-mail.

VIII. Standing

Petitioner certifies that the '675 patent is available for *inter partes* review and Petitioner is not barred or estopped from requesting *inter partes* review on the grounds herein.

IX. Conclusion

Petitioner has established a reasonable likelihood of prevailing and requests the Board institute *inter partes* review and cancel each challenged claim.

Date: July 14, 2025

Respectfully submitted,

/Erika H. Arner/

Erika H. Arner

Lead Counsel

Reg. No. 57,540

CERTIFICATE OF COMPLIANCE

The undersigned hereby certifies that the foregoing **Petition for *Inter Partes***
Review contains 12,945 words, excluding those portions identified in 37 C.F.R.
§ 42.24(a), as measured by the word-processing system used to prepare this paper.

/Erika H. Arner/
Erika H. Arner
Lead Counsel
Reg. No. 57,540

CERTIFICATE OF SERVICE

The undersigned certifies that the foregoing **Petition for *Inter Partes* Review of U.S. Patent No. 11,244,675, the associated Power of Attorney, and Exhibits 1001-1007, 1009-1012** were served on July 14, 2025, by FedEx Priority Overnight® on the correspondence address of record indicated in the Patent Office's Patent Center system for U.S. Patent No. 11,244,675.

Kevin Sprenger
DAIGNAULT IYER LLP
8229 Boone Boulevard
Suite 450
Vienna, VA 22182

Date: July 14, 2025

/Lisa C. Hines/
Lisa C. Hines
Case Manager
FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, LLP