

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

DECLARATION OF STUART J. LIPOFF

TABLE OF CONTENTS

I.	INTRODUCTION	11
II.	SUMMARY OF OPINIONS.....	12
III.	QUALIFICATIONS AND BACKGROUND	12
IV.	MATERIALS CONSIDERED	19
V.	LEGAL STANDARDS	20
A.	Level Of Ordinary Skill.....	22
B.	Prior Art.....	22
C.	Anticipation	23
D.	Obviousness.....	23
VI.	THE '675 PATENT (Ex. 1001).....	28
VII.	LEVEL OF ORDINARY SKILL IN THE ART	32
VIII.	CLAIM CONSTRUCTION	32
IX.	THE PRIOR ART	32
A.	Ocampo (Ex. 1005)	33
B.	Raitio (Ex. 1006)	38
C.	Fahlman (Ex. 1007).....	44
X.	GROUND OF UNPATENTABILITY.....	46
A.	Ground 1: Claims 1-7 Are Obvious over Raitio and Fahlman	47
1.	Independent Claim 1	47
a.	[1preamble]: “An output-content control device comprising:”	47
b.	[1a]: “a voice acquiring unit configured to acquire a voice spoken by a user;”	49

- 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;” 51
- 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;” 57
- 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” 60
- 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;” 63
- 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.” 66
 - i. Motivation to Combine Raitio and Fahlman.....68
 - ii. Expectation of Success.....69
- 2. 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.” 70

3. 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.” 73

4. 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.” 74

5. 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.” 76

6. Independent Claim 6 77

a. [6preamble]: “An output-content control method comprising:” 77

b.	[6a]: “acquiring a voice spoken by a user;”	79
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;”	80
d.	[6c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”	80
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”	80
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”	81
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.”	81
7.	Independent Claim 7	81
a.	[7preamble]: “A non-transitory storage medium that stores an output-content control program that causes a computer to execute:”	81
b.	[7a]: “acquiring a voice spoken by a user;”	82
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;”	83

d.	[7c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;”	83
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”	83
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”	84
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.”	84
B.	Ground 2: Claims 1-7 Are Obvious over Ocampo and Fahlman.....	84
1.	Independent Claim 1	84
a.	[1preamble]: “An output-content control device comprising:”	84
b.	[1a]: “a voice acquiring unit configured to acquire a voice spoken by a user;”	86
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;”	88
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;”	91
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”	95
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,”	97
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.”	102
i.	Motivation to Combine Ocampo and Fahlman	103
ii.	Expectation of Success.....	104
2.	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.”	105

3. 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.” 109
4. 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.” 110
5. 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.” 111
6. Independent Claim 6 113
 - a. [6preamble]: “An output-content control method comprising:” 113
 - b. [6a]: “acquiring a voice spoken by a user;”.....114
 - 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;” 115

- d. [6c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;” 115
- 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” 115
- 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” 116
- 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.” 116
- 7. Independent Claim 7 116
 - a. [7preamble]: “A non-transitory storage medium that stores an output-content control program that causes a computer to execute:” 116
 - b. [7a]: “acquiring a voice spoken by a user;” 117
 - 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;” 117
 - d. [7c]: “analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;” 117
 - 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”	118
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”	118
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.”	118
XI.	CONCLUSION.....	119

I. INTRODUCTION

1. I have been retained by Google LLC (“Petitioner”) as an independent expert consultant in this proceeding before the Patent Trial and Appeal Board of the United States Patent and Trademark Office.

2. I am over 18 years of age and, if I am called upon to do so, I would be willing and able to testify as to the matters set forth herein.

3. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of any related proceeding.

4. I understand that this proceeding involves U.S. Patent No. 11,244,675 (“the ’675 Patent”) (Ex. 1001). The application for the ’675 Patent was filed March 7, 2019, as U.S. Patent Application No. 16/295,034.

5. I have been asked by Petitioner to provide my opinion on whether the claims of the ’675 Patent would have been obvious to a person of ordinary skill in the art (“POSITA”) at the time of the earliest claimed priority date of the ’675 Patent. In performing my analysis, I have been asked to assume that the priority date is March 12, 2018, the date of Japanese application JP2018-044598, to which the ’675 patent claims priority. My opinions are set forth below.

6. Throughout this declaration, I refer to specific pages, figures, or line numbers of various exhibits. These citations are illustrative and are not intended to suggest that they are the only support for the propositions for which they are cited.

II. SUMMARY OF OPINIONS

7. This declaration considers claims 1-7 of the '675 Patent. Below, I set forth the opinions I have formed, the conclusions I have reached, and the bases for these opinions and conclusions. I believe the statements contained in this declaration to be true and correct to the best of my knowledge.

8. Based on my experience and knowledge of the art at the time of the earliest claimed priority date of March 12, 2018, it is my opinion that claims 1-7 of the '675 Patent would have been obvious based on the asserted grounds discussed below. A person of ordinary skill in the art ("POSITA") would have found it obvious to combine the prior art references I cite and would have been motivated to do so before the priority date.

III. QUALIFICATIONS AND BACKGROUND

9. I believe that I am well qualified to serve as a technical expert in this matter based upon my qualifications, discussed in detail below.

10. I am currently president of IP Action Partners Inc., a consulting practice that serves the telecommunications, information technology, media, electronics, and e-business industries.

11. I earned a Bachelor of Science degree in Electrical Engineering in 1968 and a second Bachelor of Science degree in Engineering Physics in 1969, both from Lehigh University. I earned a Master of Science degree in Electrical Engineering from Northeastern University in 1974, and then a Master of Business Administration degree from Suffolk University in 1983.

12. I hold a Federal Communications Commission (“FCC”) General Radiotelephone License. I also hold a Certificate in Data Processing from the Institute for the Certification of Computing Professionals (“ICCP”), which is supported by the Association for Computing Machinery (“ACM”).

13. I am also a registered professional engineer (PE) in the Commonwealth of Massachusetts and in the State of Nevada.

14. I am a fellow of the Institute of Electrical and Electronics Engineers (“IEEE”) Consumer Electronics, Communications, Computer, Circuits, and Vehicular Technology Groups. I have been a member of the IEEE Consumer Electronics Society National Board of Governors (formerly known as the Administrative Committee) since 1981, and I was Boston Chapter Chairman of the IEEE Vehicular Technology Society from 1974 to 1976. I served as the 1996-1997 President of the IEEE Consumer Electronics Society, and from 1999 to 2018 I served as Chairman of the Society’s Technical Activities and Standards Committee and as Vice President of Publications for the Society. From 2018 to 2023 I served

as Vice President of Standards and Industry Activities for the Society and currently serve on the Board of Governors as The Historian for The Society. I have also served as an Ibuka Award committee member for the IEEE's Award in the field of consumer electronics.

15. I have prepared and presented numerous papers at the IEEE and at other professional meetings. For example, in fall 2000, I served as general program chair for IEEE's Vehicular Technology Conference on advanced wireless communication technology. I have organized sessions at The International Conference on Consumer Electronics, and I was the 1984 program chairman. I conducted an eight-week IEEE-sponsored short course on Fiber Optics System Design. I received IEEE's Centennial Medal in 1984, and I received IEEE's Millennium Medal in 2000.

16. As Vice President and Standards Group Chairman for the Association of Computer Users ("ACU") from 1980 to 1983, I served as the ACU representative to the ANSI X3 Standards Group. From 1976 to 1978, I served as Chairman of the task group on user rule compliance for the FCC's Citizens Advisory Committee on Citizen's Band Radio.

17. Over the last 25 years, I have been a member of the Society of Cable Television Engineers, the Association for Computing Machinery, and The Society of Motion Picture and Television Engineers. From 2001 to 2004, I served as a

member of the USA advisory board to the National Science Museum of Israel. In 1998, I presented a short course on international product development strategies as a faculty member for Technion Institute of Management in Israel. From 2001 to 2003, I served as a member of the board or directors of The Massachusetts Future Problem Solving Program.

18. I am a named inventor on seven United States patents and have several publications on data communications in publications, including Electronics Design, Microwaves, EDN, the Proceedings of the Frequency Control Symposium, Optical Spectra, and IEEE publications.

19. During my professional career dating from 1969 to the present, I have been heavily engaged in the study, analysis, evaluation, design, and implementation of products and technology associated with consumer electronics and electronic appliances. A particular focus of my professional activities has been improving the man-machine interface including voice, speech, and speaker recognition for man-machine interactions. I also have extensive experience in studying foundation technologies and the applications supporting speech signal processing.

20. For approximately three years, from 1969 to 1972, I served as Project Engineer for Motorola's Communications Division, where I had project design responsibilities for paging and wireless communication products. Projects I worked

on while employed at Motorola included work on paging systems that included digital voice storage, voice compression, and voice synthesis. I also worked on projects that interfaced wireless data communications terminals to public safety computer systems for mobile data retrieval and data entry.

21. For approximately four years, from 1972 to 1976, I served as Section Manager for Bell & Howell Communications Company, where I also had project design responsibilities for paging and wireless communication products. The projects I supported included covert audio intelligence systems that recognized speech and activated digital voice compression recording systems. I also led projects for voice-based radio paging systems that recorded speech input, processed the speech to remove silence, processed the speech to digitally compress the speech, and store and forward the speech upon demand from DTMF or computer keyboard retrieval from the servers.

22. For 25 years from 1976 to 2001, I worked for Arthur D. Little, Inc. (ADL), where I became the Vice President and Director of Communications, Information Technology, and Electronics (CIE) and served in that role for 10 years, from 1991 to 2001. At ADL, I was responsible for the firm's global CIE practice in laboratory-based contract engineering, product development, and technology-based consulting. I was also involved in multiple pioneering efforts to identify and explore customer-to-business and business-to-business electronic

commerce and transactions information processing opportunities (e-commerce).

These projects involved technology assessment and analysis as well as developing architectures and systems to support multiple applications, and typically involved an information retrieval component.

23. While at ADL, I worked on several projects involving the combination of voice interfaces (including speech recognition and voice audio output) and information retrieval. For example, over the course of three years in the early-1990s, I worked on a project for Bolt Beranek and Newman (BB&N), where I evaluated and benchmarked technology for a voice input/output application that allowed end users (*e.g.*, travel agents) to use speech inputs to interact with airline reservation databases to retrieve information about travel reservation options, where the results were returned to the user in an audible message. This system included a natural language front-end speech-interface module with speech recognition that used pre-defined recognition grammars to convert the end user's speech into structured commands supported by an airline reservation system. As another example, over the course of three years in the mid-1990s, I worked on a project for Texas Instruments that applied a speech-recognition interface for a variety of applications that retrieved information from database servers.

24. Other projects that I worked on at various points in my 25 years at ADL and afterwards that involved speech recognition technologies included the following.

25. Over the course of three years in the early 1990s, I worked on a voice-interface project developing spoken digit telephone number recognition and voiceprint matching for Sprint's long distance alternative access telephone services.

26. Over the course of a year in the late 1980s, I worked on a voice interface project evaluating the processing power needed to perform various voice recognition applications by Rockwell Semiconductor's signal processing technology.

27. Over the course of 15 years starting in the early 1980's, I worked on a project for the United States Postal Service (USPS), where we developed a real-time automated postal teller system that served as an interface between end-users and the USPS's information systems. This system included voice prompts for the vision impaired.

28. I also have extensive experience in public and private network wired and wireless voice telecommunications while employed by Motorola, Bell & Howell, and Arthur D Little, and while self-employed. In the course of these telecommunications projects ranging from 1969 to the present, I have encountered

a number of applications where audio input and voice are used to activate devices, for example for the purpose of saving battery power by entering into low power, so-called “sleep” modes. These projects have involved the design of cellular telecommunications systems that implement industry standard means of entering lower power modes in the absence of voice.

29. My curriculum vitae are provided as Exhibit 1004.

IV. MATERIALS CONSIDERED

30. In forming my opinions, I have reviewed the following documents, as well as other documents cited throughout this declaration:

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
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”)
1011	U.S. Patent Application No. 11/586,862
1012	U.S. Patent Application No. 11/638,251

31. I have also relied on my education, experience, research, training, and knowledge in the relevant art, and my understanding of legal principles described in this declaration.

32. All of the opinions contained in this declaration are based on the documents I reviewed and my knowledge and professional judgment. My opinions have also been guided by my understanding of how a POSITA would have understood the claims of the '675 Patent at the time of the earliest claimed priority date.

33. I reserve the right to supplement and amend any of my opinions in this declaration based on documents, testimony, and other information that becomes available to me after the date of this declaration.

V. LEGAL STANDARDS

34. I am an engineer and not a lawyer. My understanding of the legal standards to apply in reaching the conclusions in this declaration is based on discussions with counsel for Petitioner, my experience applying similar standards in other patent-related matters, and my reading of the documents submitted in this proceeding. I have applied these legal standards in preparing this declaration.

35. I have been informed that there are two ways in which prior art may render a patent claim unpatentable. First, I have been informed that the prior art can “anticipate” a claim. Second, I have been informed that the prior art can render

a claim “obvious” to a POSITA. I understand that a claim is patentable if it was not anticipated and would not have been rendered obvious by the prior art at the effective filing date of the patent.

36. I have been informed that a dependent claim is a patent claim that refers back to another patent claim. I have been informed that a dependent claim includes all of the limitations of the claim to which it refers plus its own limitation(s).

37. I have been asked to provide my opinions as to whether the cited prior art discloses or renders obvious claims 1-7 of the '675 Patent from the perspective of a POSITA at the time of the earliest claimed priority date of March 12, 2018, as described in more detail below.

38. I have been informed that in IPR proceedings, such as this one, the party challenging the patent bears the burden of proving unpatentability by a preponderance of the evidence. I understand that a preponderance of the evidence means “more likely than not.”

39. For purposes of this declaration, I have been asked to provide my opinions on issues regarding unpatentability. I have been informed of the following legal standards, which I have applied in forming my opinions.

A. Level Of Ordinary Skill

40. I have been informed that a POSITA is determined by considering several factors, including the (i) type of problems encountered in the art; (ii) prior art solutions to those problems; (iii) rapidity with which innovations are made; (iv) sophistication of the technology; and (v) educational level of active workers in the field.

41. I have been instructed to assume that a POSITA is not a specific real individual, but rather a hypothetical individual having the qualities reflected by the factors discussed above. A POSITA is assumed to be person of ordinary creativity familiar with the prior art as of the priority date of the patent at issue.

B. Prior Art

42. I have been advised and understand that the information used to evaluate whether an invention was new and not obvious when made is generally referred to as “prior art.” I understand that in an IPR proceeding, prior art includes patents and printed publications that existed before the earliest claimed priority date or the earliest filing date of the patent (which I have been informed is also called the “effective filing date”). I have been informed and understand that a patent or published patent application is prior art if it was filed before the earliest filing date of the claimed invention and that a printed publication is prior art if it was publicly available before the earliest filing date.

C. Anticipation

43. I have been informed that under 35 U.S.C. § 102, a patent claim is unpatentable for anticipation if the claimed subject matter was patented or described in a printed publication before the effective filing date of the claimed invention. I have been informed that this is referred to as unpatentability by anticipation. I have been informed that a patent claim is anticipated under § 102 if a single prior art reference discloses all the limitations of the claimed invention. I understand that limitations may be expressed or inherent such that the limitation is essential to the prior art.

D. Obviousness

44. I have been informed that for obviousness under 35 U.S.C. § 103, a patent claim is unpatentable if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious to a POSITA to which said subject matter pertains at the time the invention was made. I have been informed that this is referred to as unpatentability by obviousness.

45. I have been informed that an obviousness analysis includes the following considerations:

- a. Determining the scope and content of the prior art;

- b. Ascertaining the differences between the prior art and the claims at issue;
- c. Resolving the level of ordinary skill in the pertinent art; and
- d. Considering evidence of secondary indicia of nonobviousness (if available).

46. I have been informed that the relevant time for considering whether a claim would have been obvious to a POSITA is the time of invention. For my obviousness analysis, counsel for Petitioner instructed me to assume that the date of invention for the challenged claims is March 12, 2018. My opinions would not change if I assumed another, e.g., later, date of invention.

47. I have been informed that a reference may be modified or combined with other references or with a POSITA's own knowledge if the person would have found the modification or combination obvious. I have also been informed that a POSITA is presumed to know all the relevant prior art, and the obviousness analysis may take into account the inferences and creative steps that a POSITA would employ.

48. I have been informed that an obviousness determination must be made from the perspective of a POSITA. I have also been informed that there is no requirement that the prior art contain an express suggestion to combine known elements to achieve the claimed invention, and that a suggestion to combine known

elements to achieve the claimed invention may come from the prior art as a whole or individually. Also, the obviousness analysis may rely on the inferences and creative steps a POSITA would employ, as filtered through his or her knowledge as of the priority date. But I understand that obviousness grounds cannot be sustained by mere conclusory statements and must include some articulated reasoning and rationale to support a legal conclusion of obviousness.

49. In determining whether a prior art reference could have been combined with another prior art reference or other information known to a POSITA, I have been informed that the following principles may be considered:

- a. A combination of familiar elements according to known methods is likely to be obvious if it yields predictable results;
- b. The substitution of one known element for another is likely to be obvious if it yields predictable results;
- c. The use of a known technique to improve similar items or methods in the same way is likely to be obvious if it yields predictable results;
- d. The application of a known technique to a prior art reference that is ready for improvement to yield predictable results;

- e. Any need or problem known in the field and addressed by the reference can provide a reason for combining the elements in the manner claimed;
- f. A person of ordinary skill often will be able to fit the teachings of multiple references together like a puzzle; and
- g. The proper analysis of obviousness requires a determination of whether a POSITA would have a “reasonable expectation of success”—but not “absolute predictability” of success—in achieving the claimed invention by combining prior art references.

50. I have been informed that, when a work is available in one field, design alternatives and other market forces can prompt variations of it, either in the same field or in another. I have been informed that if a POSITA could have implemented a predictable variation and would have seen the benefit of doing so, that variation is likely to have been obvious. I have been informed that, in many fields, such as the mechanical or electrical arts, market demand—not scientific literature—may drive design trends. I have been informed that, when there was a design need or market pressure and there are a finite number of predictable solutions, a POSITA would have had a good reason to pursue those known options.

51. I have been informed that the law permits the application of “common sense” in examining whether a claimed invention would have been obvious to a POSITA. For example, I have been informed that combining familiar elements according to known methods and in a predictable way may suggest obviousness when such a combination would yield nothing more than predictable results. I understand, however, that a claim is not obvious merely because every claim element is disclosed in the prior art. A party asserting obviousness must provide a specific motivation to combine or modify the references as recited in the claims and explain why one skilled in the art would have reasonably expected to succeed in doing so.

52. I have been informed that there is no rigid rule that a reference or combination of references must contain a “teaching, suggestion, or motivation” to combine references. But I also understand that the “teaching, suggestion, or motivation” test can be a useful guide in establishing a rationale for combining elements of the prior art. I have been informed that this test poses the question as to whether there is an express or implied teaching, suggestion, or motivation to combine prior art elements in a way that results in the claimed invention, and that it helps to counter the use of hindsight, which is impermissible. Likewise, if a prior art reference “teaches away” from a potential prior art combination, then a motivation to combine may not exist.

53. I am not aware of any evidence of secondary considerations, such as unexpected results, industry skepticism, long-felt unresolved need, commercial success, praise by others, or copying that would alter my opinions set forth below.

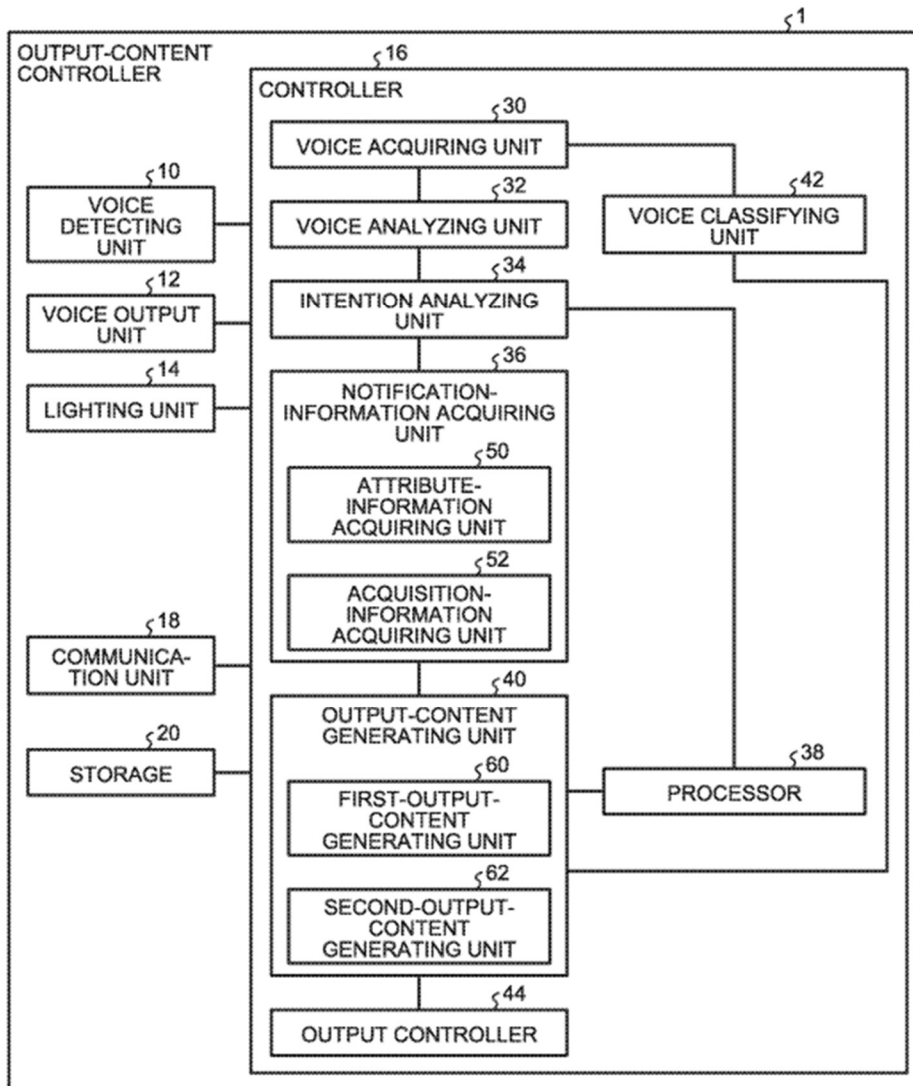
54. I have been informed that, in an obviousness analysis, prior art must be analogous art to the patent being considered. I have been informed that a prior art reference is considered to be analogous, or in the same field of art, if the reference is either (1) in the same field of endeavor as the challenged patent, regardless of the problems the challenged patent and the prior art address, or (2) reasonably pertinent to the particular problem being solved by the challenged patent.

VI. THE '675 PATENT (Ex. 1001)

55. The '675 patent was filed on March 7, 2019 issued on February 8, 2022 and claims priority to Japanese Patent Application No. JP 2018-044598 filed March 12, 2018. Ex. 1001, cover page (22), (30), (45).

56. The '675 patent discloses devices and methods for 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.

FIG.2

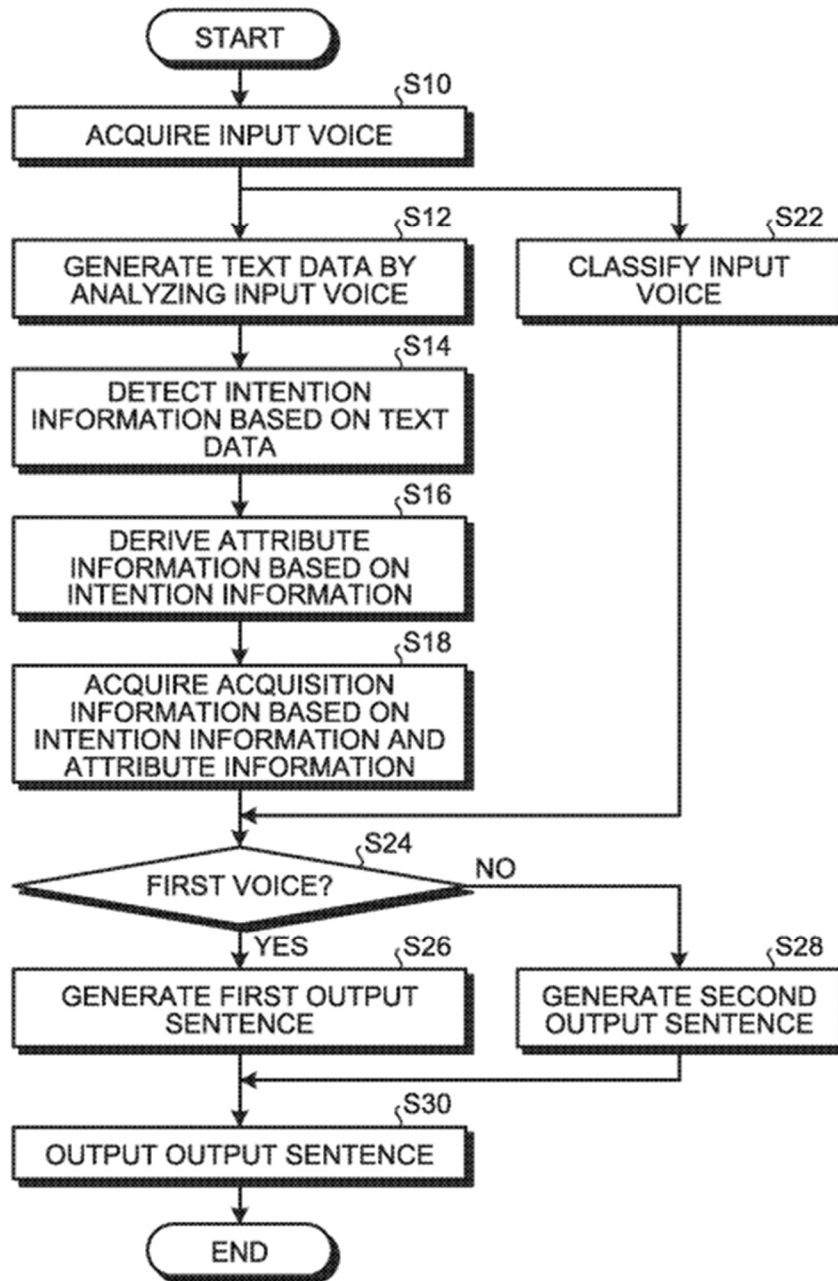


Ex. 1001, FIG. 2.

57. 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 processes 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.

VII. LEVEL OF ORDINARY SKILL IN THE ART

58. In my opinion, 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.

VIII. CLAIM CONSTRUCTION

59. I understand that, in this proceeding, the claims are construed according to their ordinary and customary meaning, in light of the specification and prosecution history, as understood by a POSITA at the time of the invention. I understand that this is the same claim construction approach used in district court litigation.

60. In my opinion none of the claim terms in the ’675 require an explicit construction.

IX. THE PRIOR ART

61. As explained in Section X below, it is my opinion that claims 1-7 would have been obvious over combinations of Raitio, Ocampo, and Fahlman, each of which I have been instructed to assume is prior art.

A. Ocampo (Ex. 1005)

62. Ocampo, filed November 1, 2016, and published May 3, 2018.

Ocampo, cover (22), (43). I understand that Ocampo is prior art to the '675 patent.

63. 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.

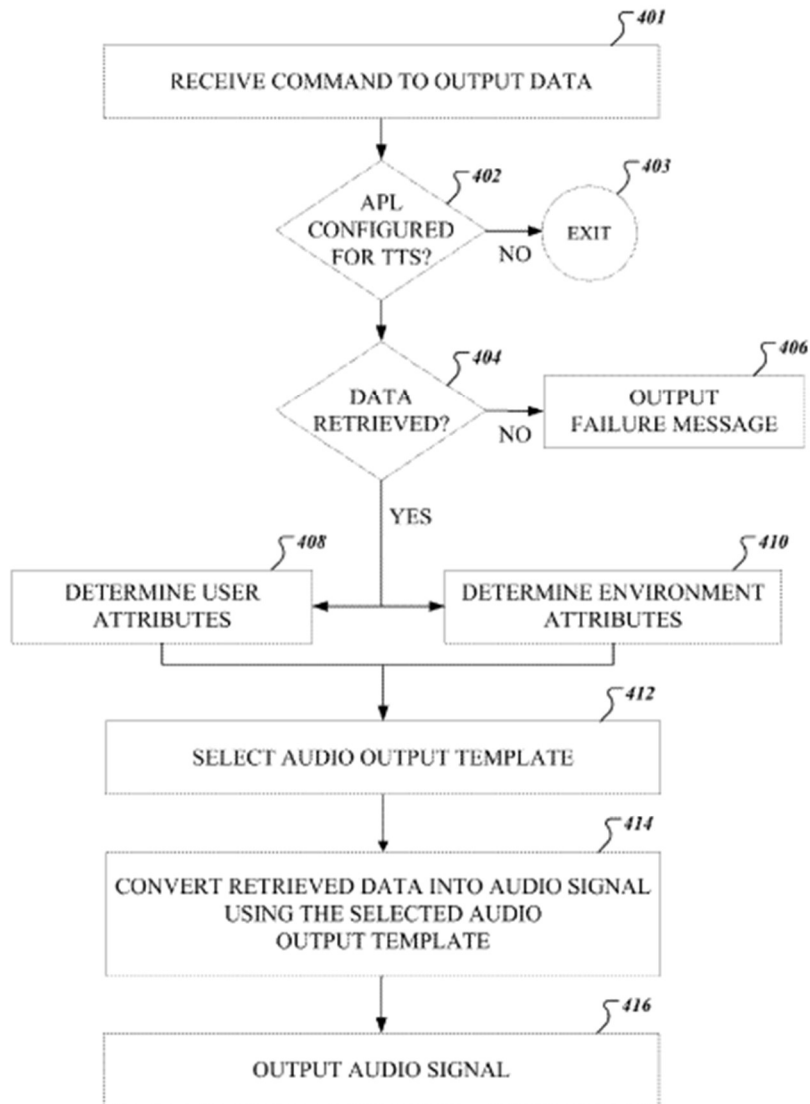
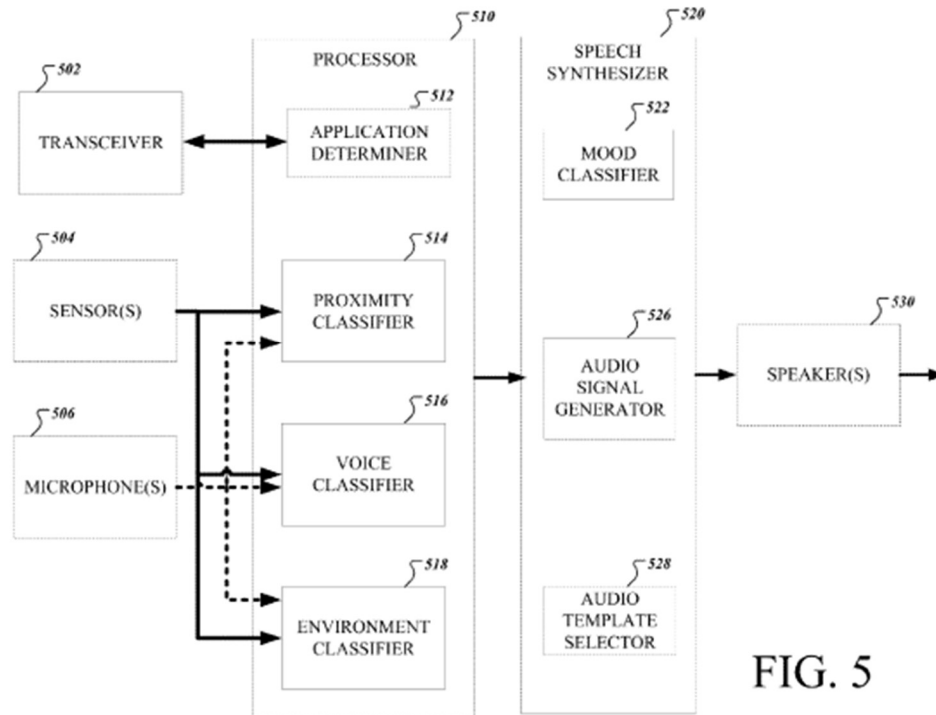


FIG. 4

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.

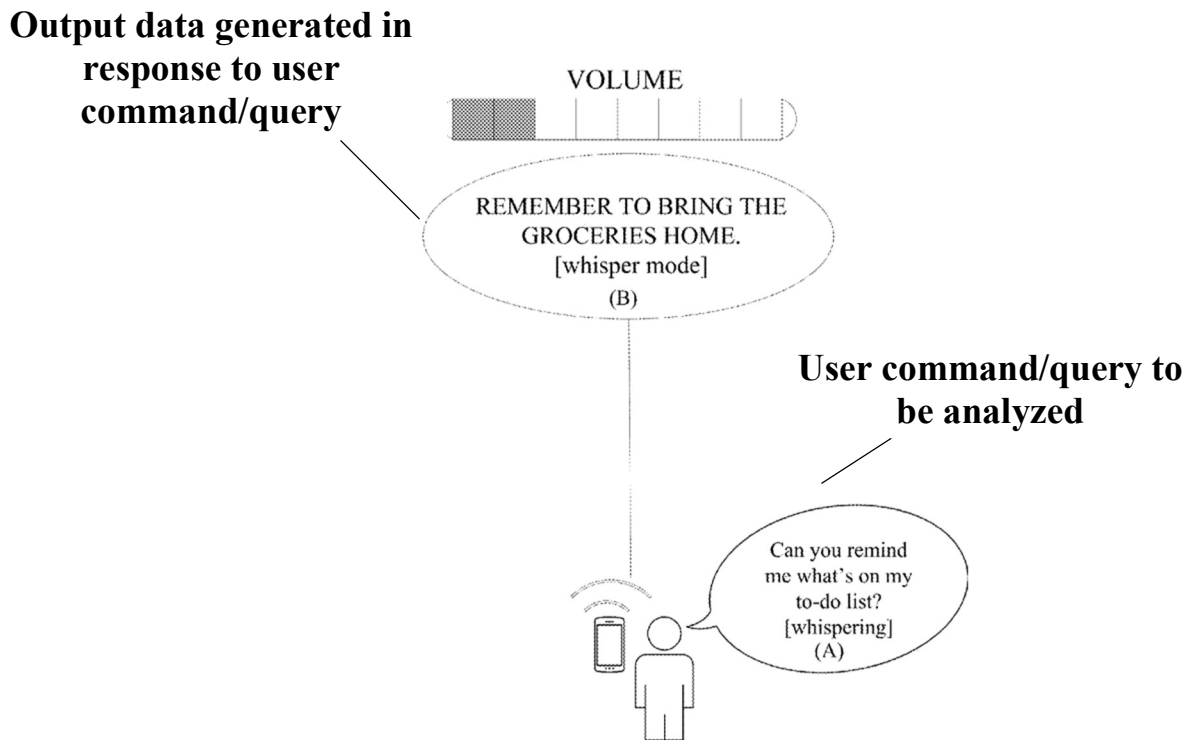


Ocampo, FIG. 5.

64. 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).

65. 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.”).

66. 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. 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.

B. Raitio (Ex. 1006)

67. Raitio was filed on September 15, 2016, and published on December 14, 2017. Raitio, cover page (22), (45). I understand that Raitio is prior art to the '675 patent.

68. 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.

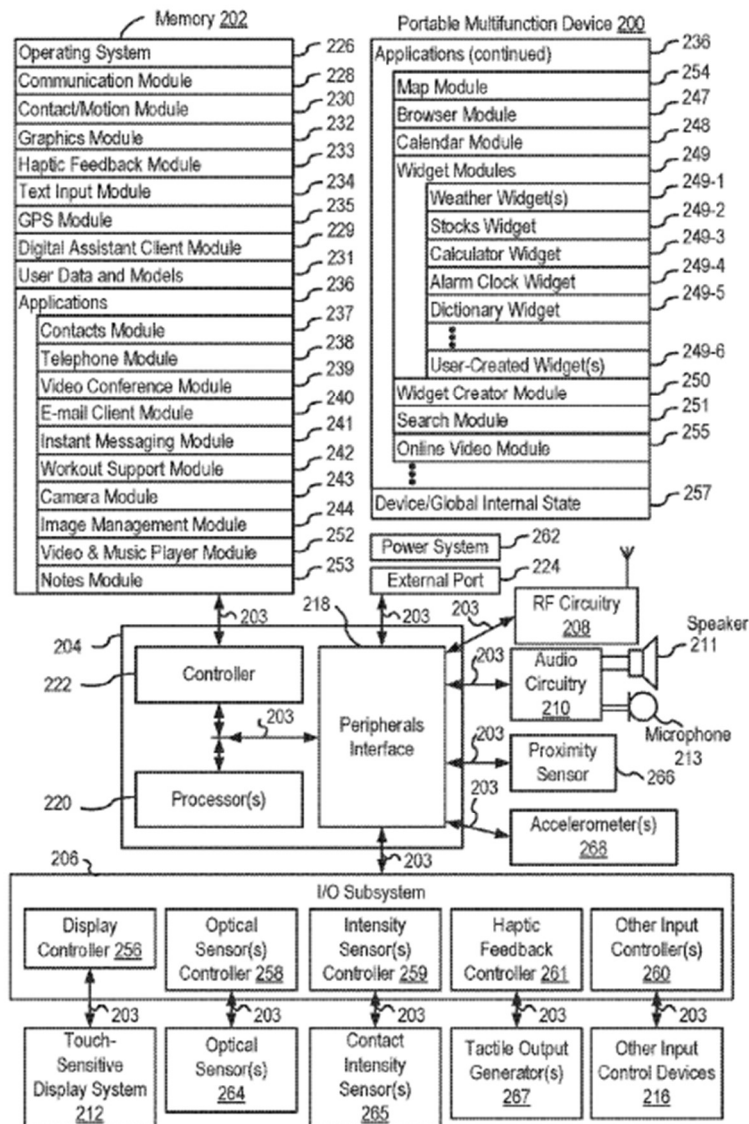


FIG. 2A

Raitio, FIG. 2A. Figure 8A below illustrates additional modules, including a “whispered speech determination module” of the digital assistant. Raitio, [0020], [0239]-[00240], FIG. 8A.

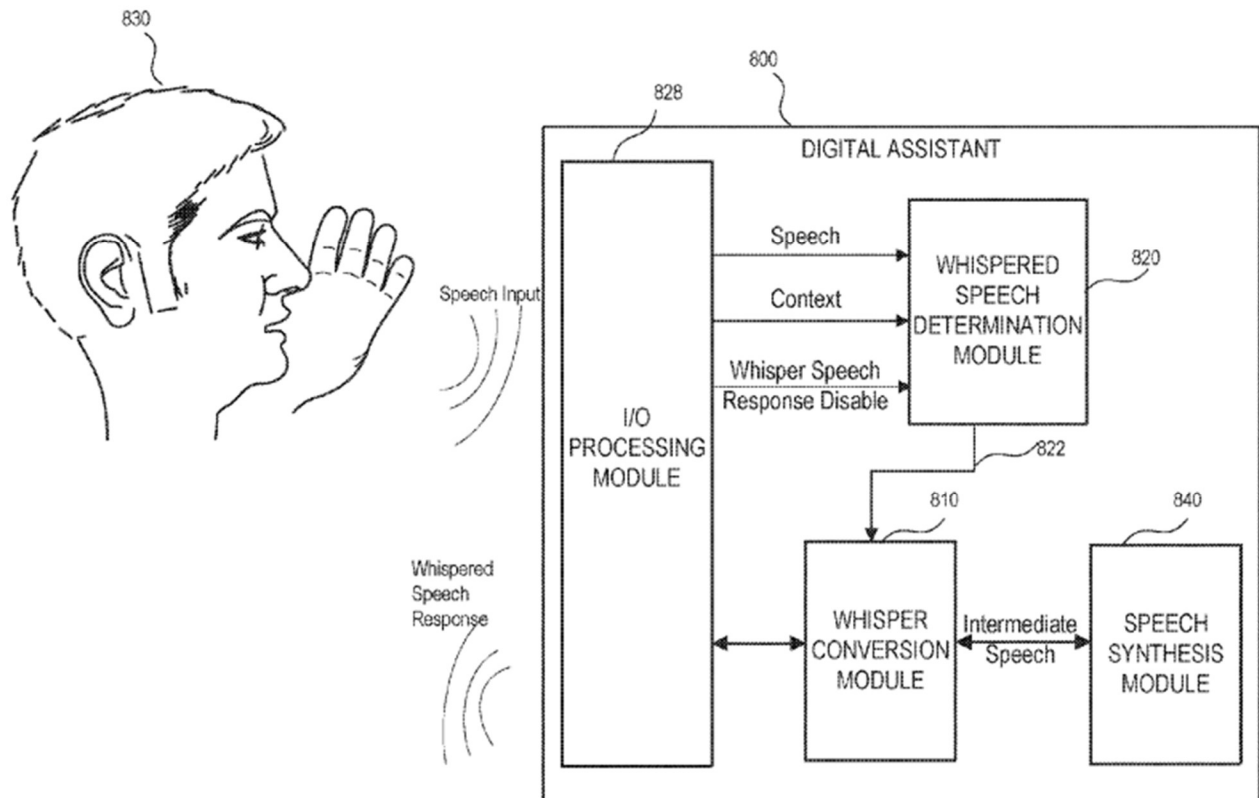


FIG. 8A

Raitio, FIG. 8A.

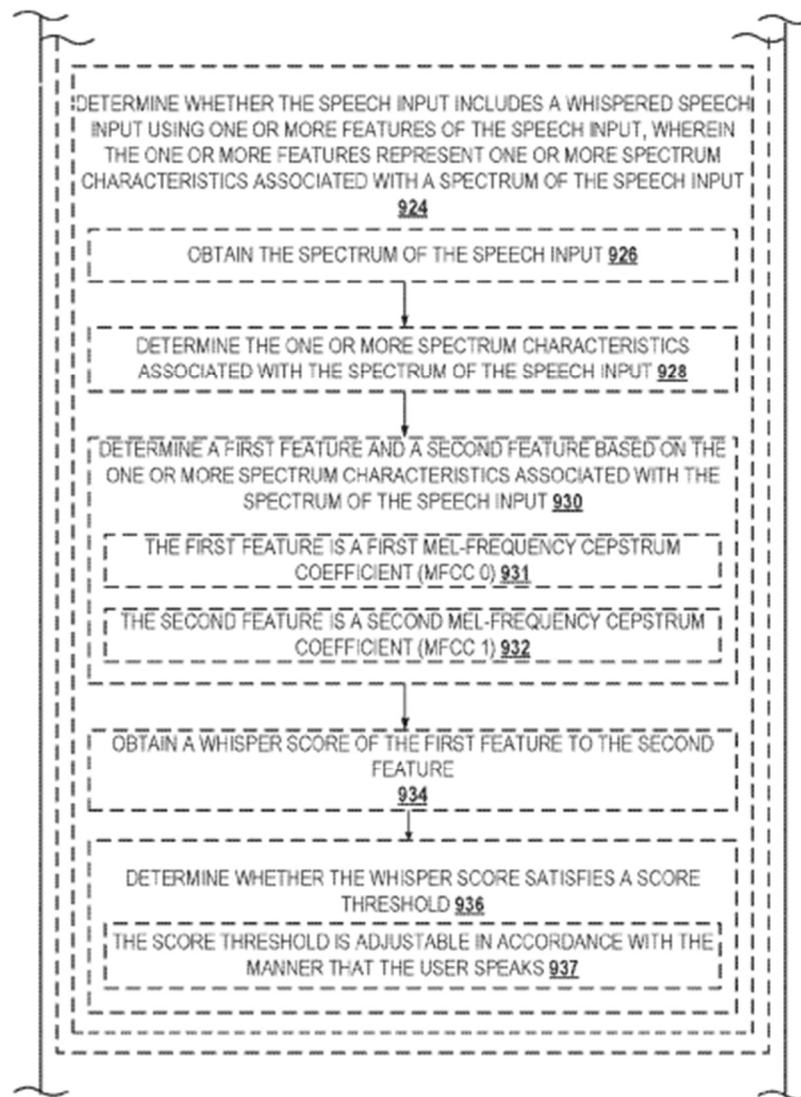
69. Like the '675 patent, Raitio's device "interprets natural language input ... to infer user intent, and performs actions based on the inferred user intent."

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").

70. 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 users 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.



Raitio, FIG. 9B.

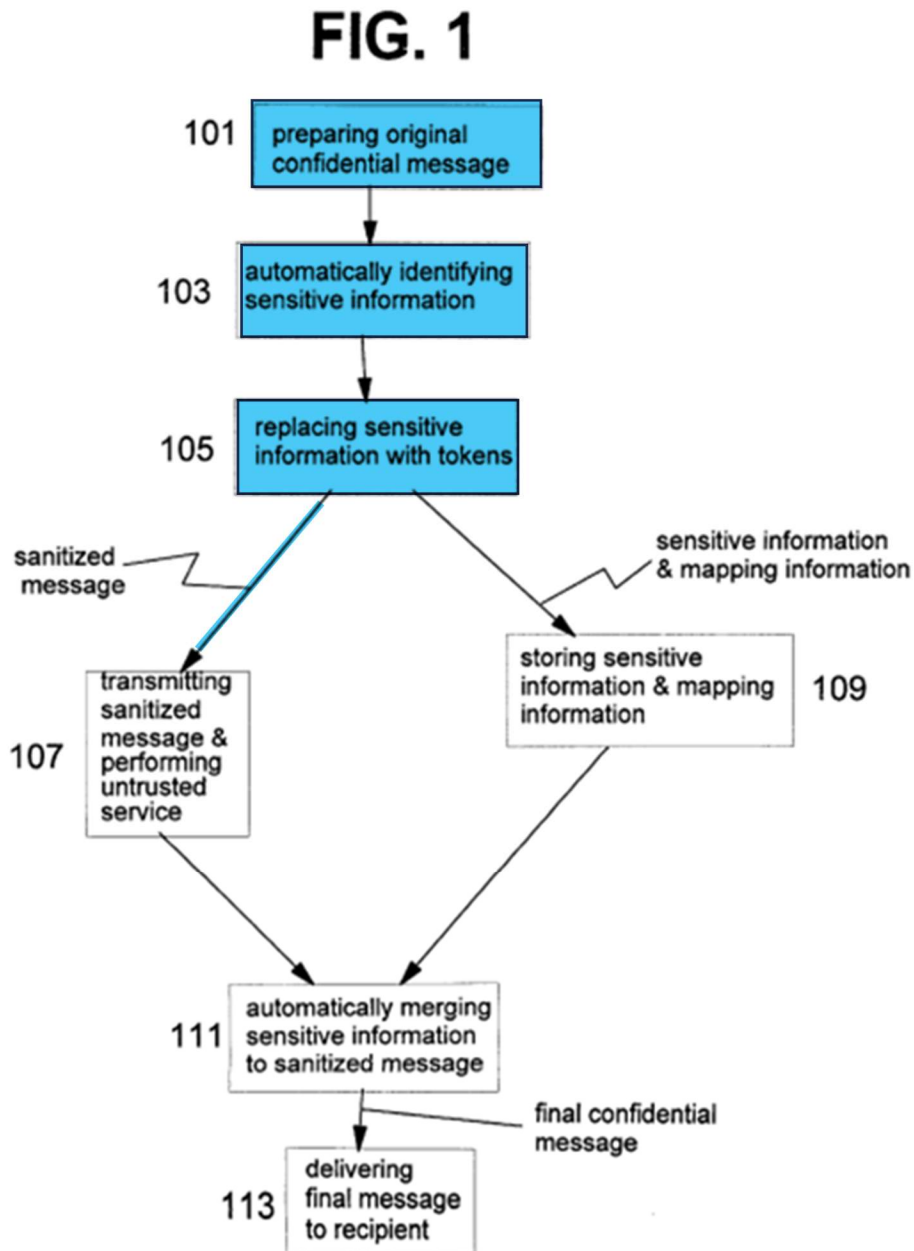
71. In my opinion, Raitio is analogous art to the '675 patent because it is in the same field of endeavor as the '675 patent: speech processing. *Compare* Raitio, Abstract, *with* Ex. 1001, Abstract. It is also my opinion that Raitio is directed to the same problem as the '675 patent: analyzing voice input commands

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

C. Fahlman (Ex. 1007)

72. Fahlman was filed on November 7, 1997, and issued September 28, 1999. Fahlman, cover page (22), (45). I understand that Fahlman is prior art to the '675 patent.

73. 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).

74. 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, it is my opinion that 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.

X. GROUNDS OF UNPATENTABILITY

75. Based on my review of the materials set forth above, including my application of the knowledge of a POSITA, it is my opinion that claims 1-7 of the ’675 Patent would have been obvious to one of ordinary skill in the art as of March 12, 2018.

76. In particular, it is my opinion that claims 1-7 of the '675 Patent would have been obvious to a POSITA based on the following grounds:

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

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

1. Independent Claim 1

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

77. 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. 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.

78. 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.

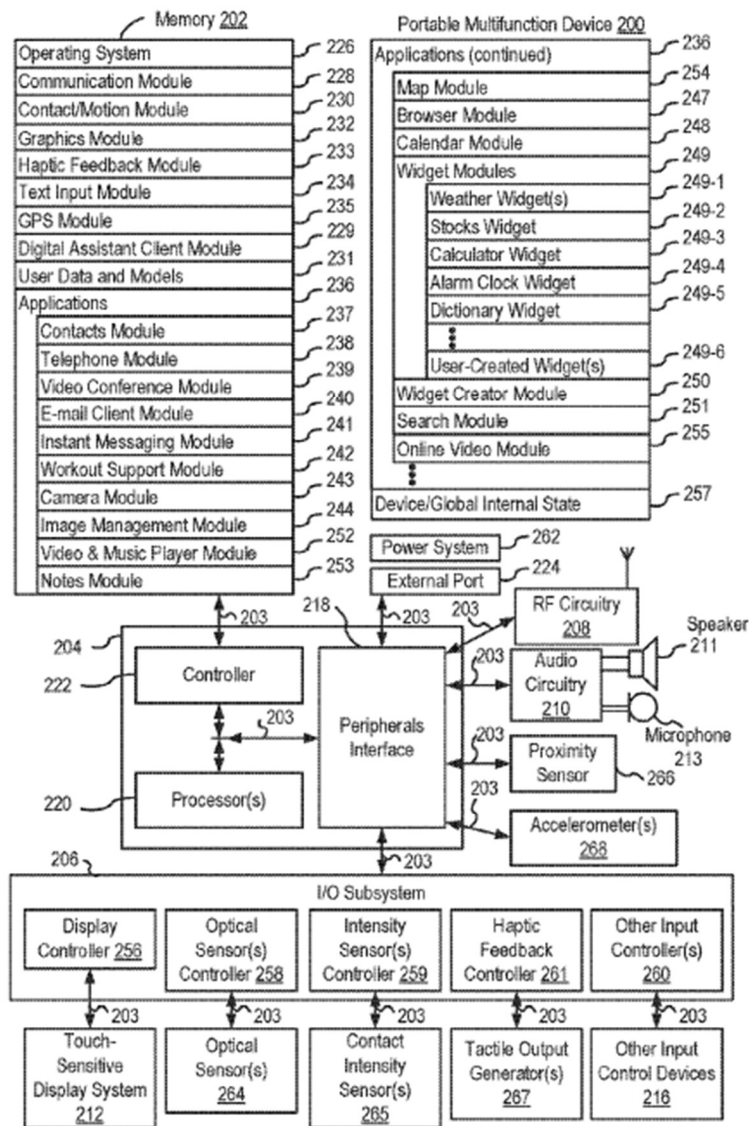


FIG. 2A

Raitio, FIG. 2A. Figure 8A below illustrates a block diagram of the functions of the digital assistant. Raitio, [0020], [0239]-[0240], FIG. 8A.

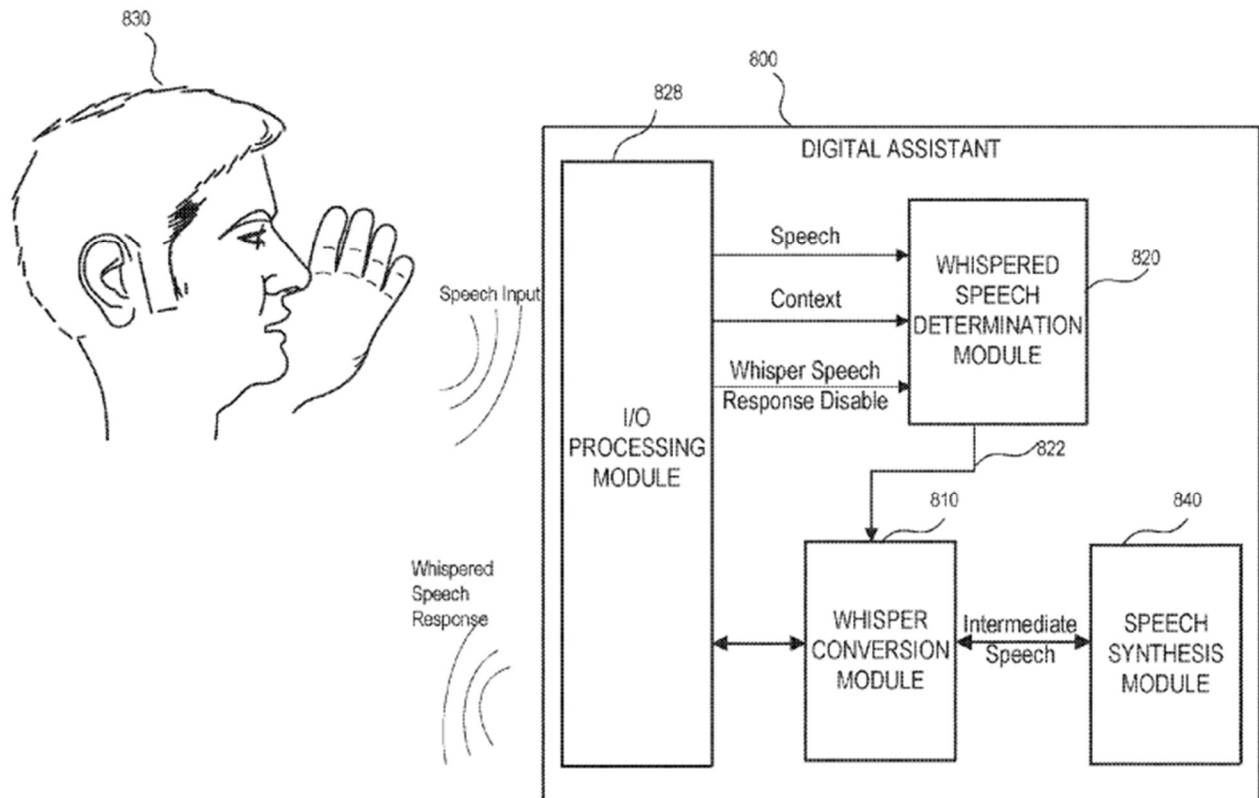


FIG. 8A

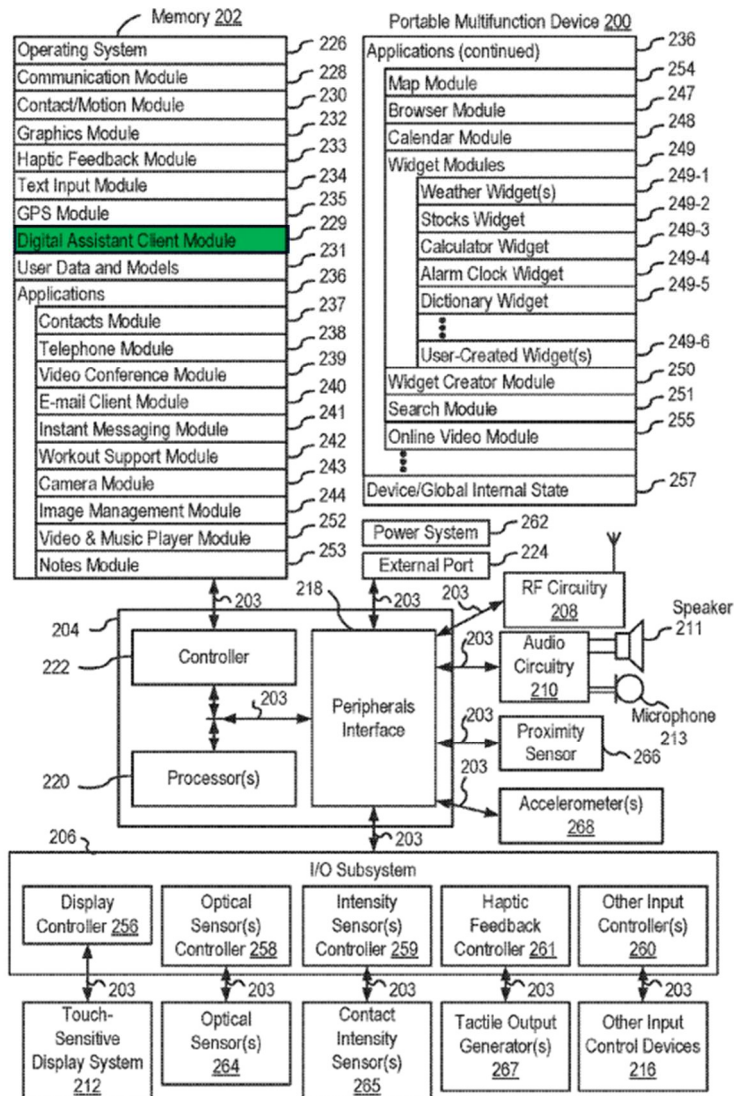
Raitio, FIG. 8A.

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

79. 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)*,”¹ text input, touch input,

¹ All emphases have been added, unless otherwise noted.

and/or gestural input through various user interfaces (e.g., microphone 213, accelerometer(s) 268, touch-sensitive display system 212, optical sensor(s) 229, other input control devices 216, etc.) 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). The module and its associated “applications (or sets of instructions)” are stored in the device’s memory (202/702). Raitio, [0066], [0207], FIG. 2, 7A-7B. This is consistent with 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.



Raitio, FIG. 2A (annotated).

- 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;”

80. 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.

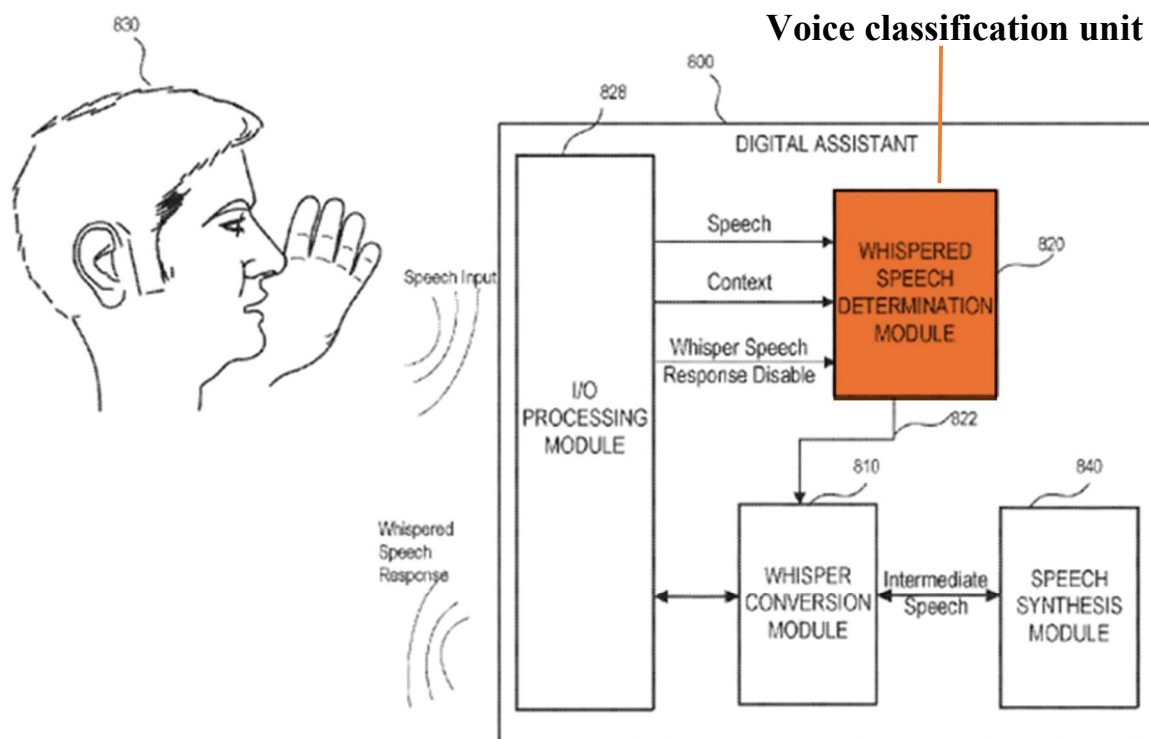


FIG. 8A

Raitio, FIG. 8A (annotated).

81. 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.”).

82. 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]. In my opinion, 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. 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.

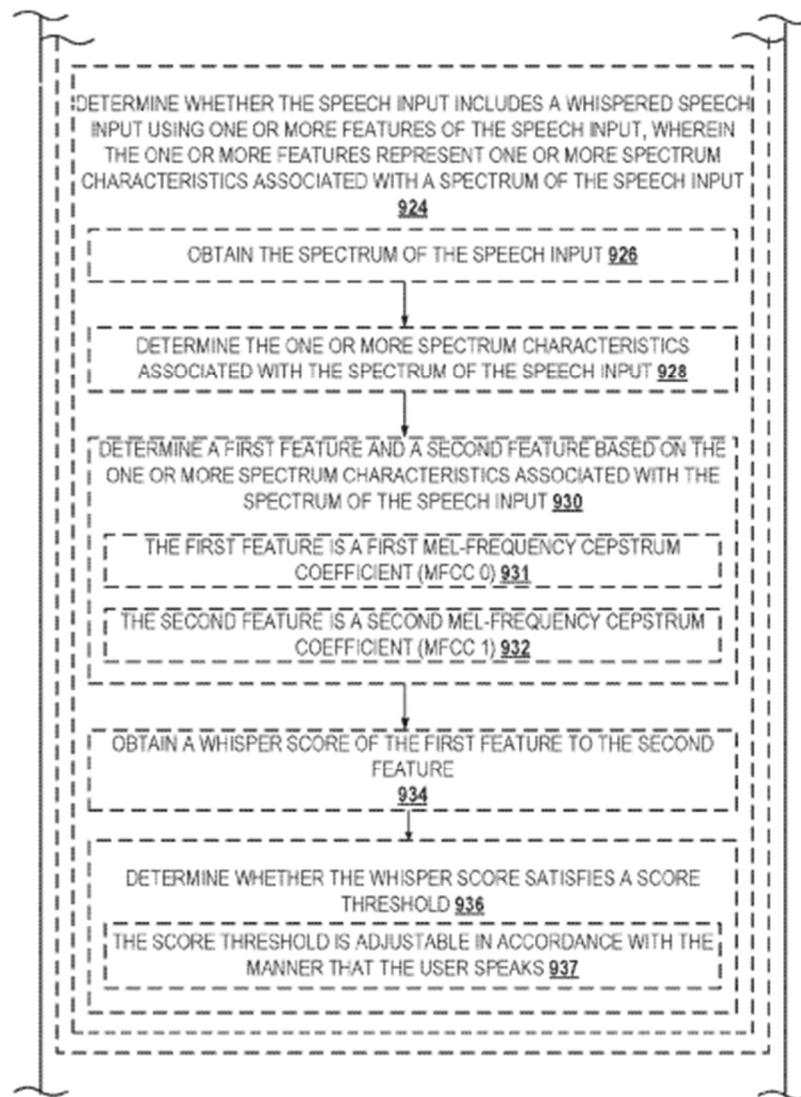


FIG. 9B

Raitio, FIG. 9B.

83. 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.

84. Furthermore, Raitio discloses that its device may include “one or more proximity sensors” Raitio, [0063]. 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. *See e.g.*, 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.). Raitio also incorporates by references the entire teachings of U.S. Application No. 11/586,862, “Automated Response To And Sensing Of User Activity In Portable Devices” and U.S. Application No. 11/638,251, “Methods And Systems For Automatic Configuration Of Peripherals,” both of which utilize proximity sensors to determine how close a user’s head is to a mobile device and uses the distance calculation to deduce volume. Ratio, [0063]; Ex. 1011, [0087], [0090]; Ex. 1012, [0094]-[0095].

85. 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

- 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;”**

86. 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; *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”).

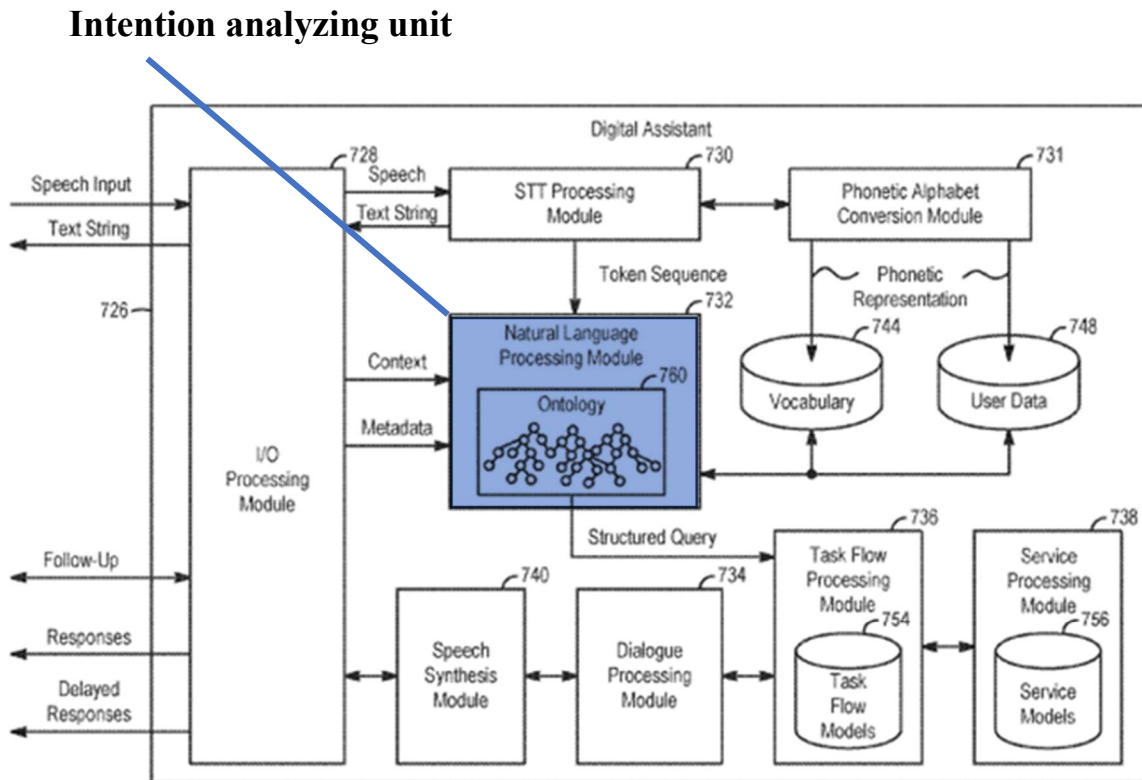


FIG. 7B

Raitio, FIG. 7B (annotated).

87. 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].

88. 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].

89. 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.)”).

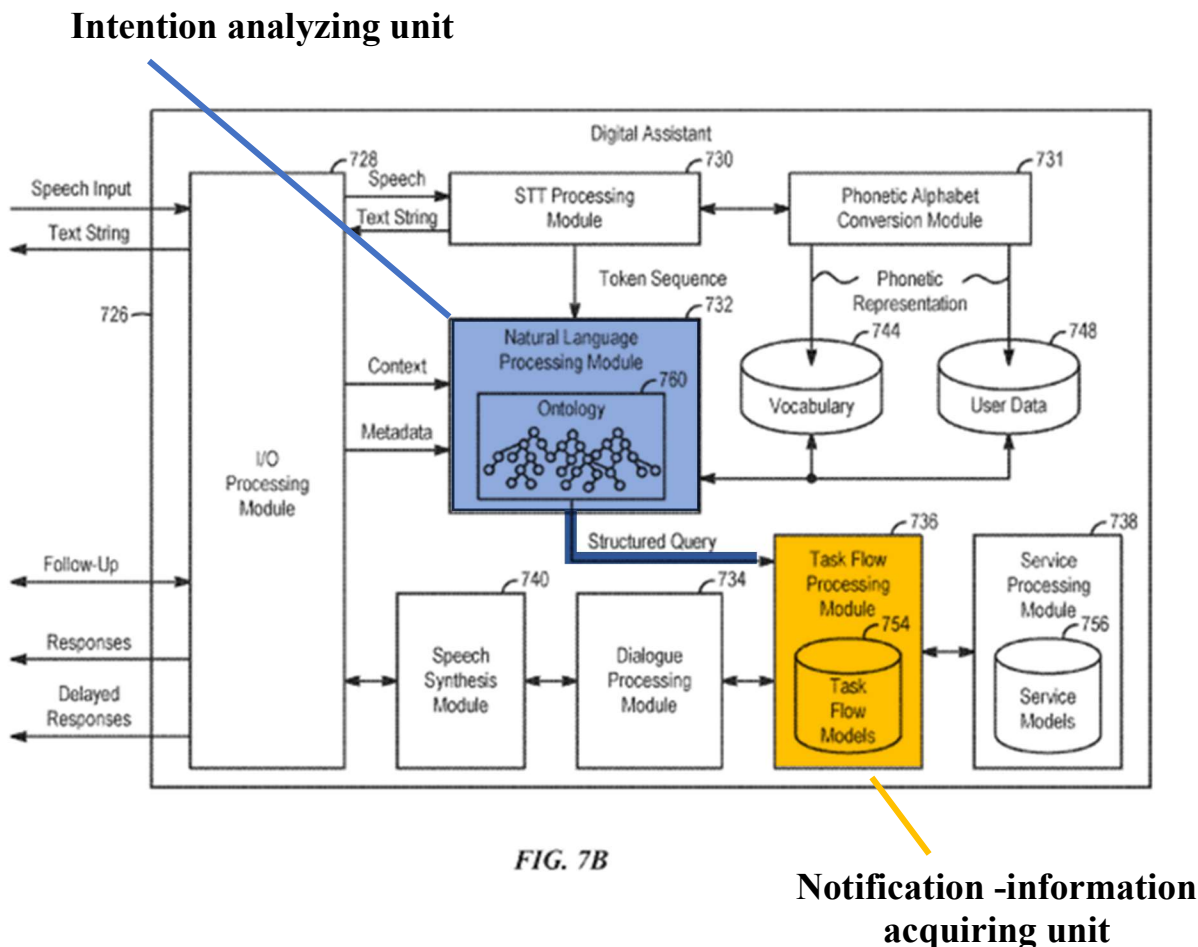
- e. **[1d]: “a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and”**

90. 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. The acquired information is subsequently presented to the user. Raitio, [0229], [0230], [0232], [0235], FIG. 7B.

91. As I explain in [1c], above, Raito’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).

92. 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]; *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* [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,”**

93. 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 (output-content generating unit). Raitio, [0235]-[0236]. 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 that digital assistant 800 implements the speech synthesis module).

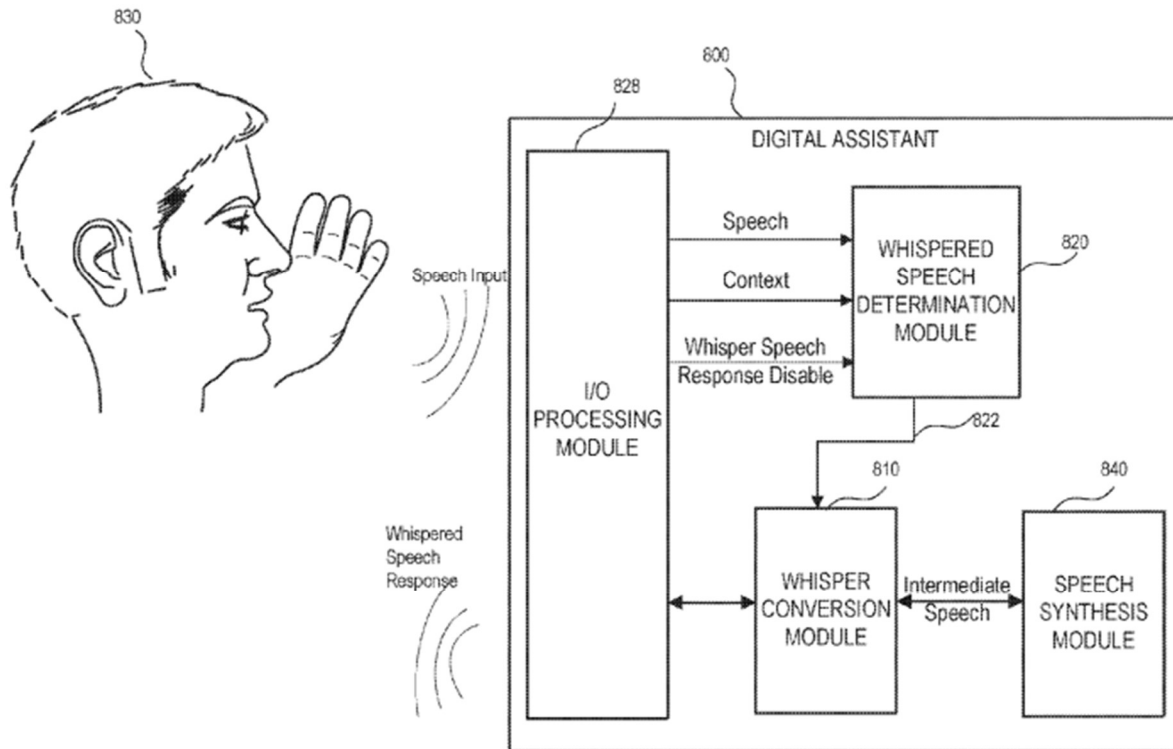


FIG. 8A

Raito, FIG. 8A.

94. As I explain 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]. Fahlman similarly discloses a system and method for generating an output message (sentence), as I explain in further detail in [1f], below. 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.”**

95. In my opinion, Raitio-Fahlman renders obvious [1f].

96. First, 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].

97. As I explain for [1e], above, 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]. And as explained for [1b], above, “[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].

98. 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.

99. 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

100. In my opinion, 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]. For example, a POSITA would have appreciated that in addition to whispering a generated response (as described above), sanitizing the response to remove any private (sensitive) information would have further enhanced the user’s experience.

101. Thus, it is also my opinion that 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.

ii. Expectation of Success

102. It is also my opinion that 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.

103. 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].

104. 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 of algorithms" may be used "for replacing the identified sensitive terms with standard tokens, thereby creating a sanitized message." Fahlman, 6:35-38.

105. Thus, a POSITA would have had a reasonable expectation of success in implementing Fahlman's word replacement algorithms in Raito's assistant

because it would have merely required the integration of computer algorithms and routine skill to do so.

2. **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.”**

106. In my opinion, Raitio discloses, and the Raitio-Fahlman combination renders obvious the additional limitations of claim 2.

107. First, Ratio discloses or Ratio-Fahlman renders obvious “*the notification information includes type information indicating a type of the content information.*” As I explain for [1d], above, 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]. In my opinion, the content information in the output also corresponds to type 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 is consistent with 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, 16:17-21, FIGS. 4-5. Thus, it is my opinion that Raitio discloses the additional limitations of claim 2.

108. I have also been asked to consider whether Fahlman renders obvious the additional limitations of claim 2. In my opinion, it does. Like Raitio, Fahlman 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).

109. Accordingly, it is also my opinion that a POSITA would have been motivated to modify the notification information of Raitio 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 I set forth in [1f], above.

110. Second, Ratio-Fahlman renders obvious *“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.”*

111. As I explain for [1f], above, 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.

112. 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.

113. 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 [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.

- 3. 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."**

114. In my opinion, the Raitio-Fahlman combination renders claim 3 obvious.

115. 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.

116. 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.

4. **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.”**

117. In my opinion, the Raitio-Fahlman combination 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.

118. Fahlman's 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.

119. 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.

5. 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 voice² when the voice spoken by the user is a whisper.”

120. It is my opinion that Raitio discloses the additional limitations of claim 5 because as I explain in [1b], above, 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.

121. 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).

² Because this is the first use of “the predetermined voice,” I understand “the predetermined voice” to mean “a predetermined voice.”

122. This is also consistent with the '675 patent which also uses spectrum characteristics to classify a voice as whispered voice, and 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, 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.”).

6. Independent Claim 6

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

123. 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.



Raitio, FIG. 9A.

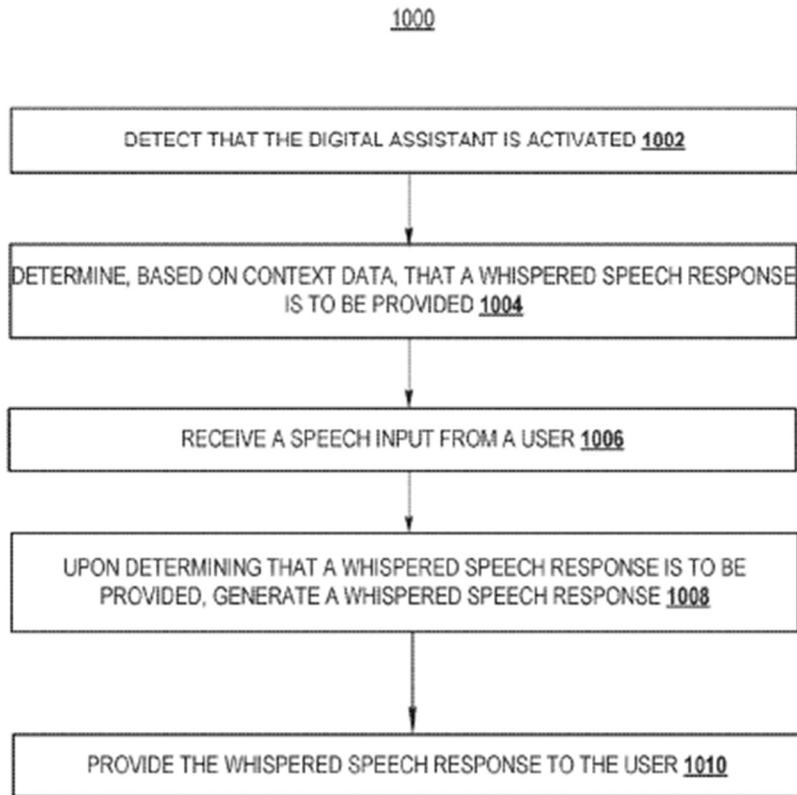


FIG. 10

Raitio, FIG. 10.

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

124. Claim element [6a] is substantively similar to claim element [1a] above; thus, it is my opinion that Raitio discloses [6a] for reasons discussed for [1a].

- 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;”**

125. Claim element [6b] is substantively similar to claim element [1b] above; thus, it is my opinion that Raitio discloses [6b] for reasons discussed for [1b].

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

126. Claim element [6c] is substantively similar to claim element [1c] above; thus, it is my opinion that Raitio discloses [6c] for reasons discussed for [1c].

- 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”**

127. Claim element [6d] is substantively similar to claim element [1d] above; thus, it is my opinion that Raitio discloses [6d] for reasons discussed for [1d].

- 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”**

128. Claim element [6e] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Raitio-Fahlman combination renders obvious [6e] for reasons discussed for [1e]-[1f].

- 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.”**

129. Claim element [6f] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Raitio-Fahlman combination renders obvious [6f] for reasons discussed for [1e]-[1f].

7. Independent Claim 7

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

130. 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], below. Raitio, [0006], [0202], [0207]-[0208], FIG. 7A.

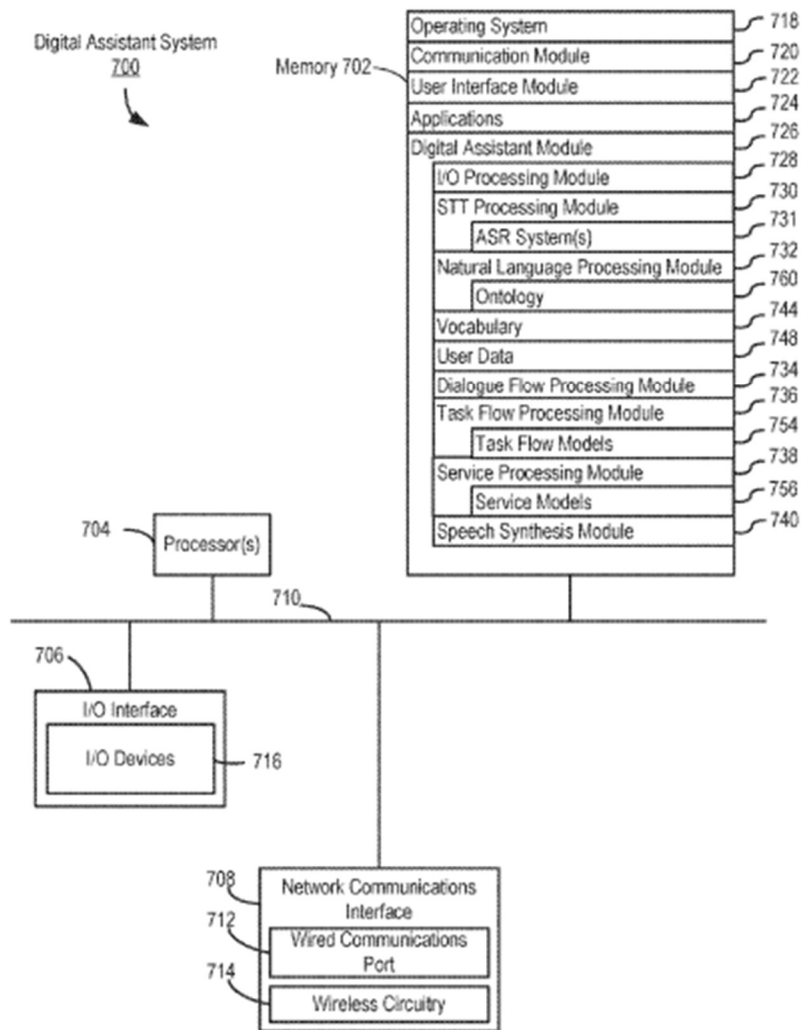


FIG. 7A

Raitio, FIG. 7A.

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

131. Claim element [7a] is substantively similar to claim element [1a] above; thus, it is my opinion that Raitio discloses [7a] for reasons discussed for [1a].

- 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;”**

132. Claim element [7b] is substantively similar to claim element [1b] above; thus, it is my opinion that Raitio discloses [7b] for reasons discussed for [1b].

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

133. Claim element [7c] is substantively similar to claim element [1c] above; thus, it is my opinion that Raitio discloses [6c] for reasons discussed for [1c].

- 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”**

134. Claim element [7d] is substantively similar to claim element [1d] above; thus, it is my opinion that Raitio discloses [7d] for reasons discussed for [1d].

- 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”**

135. Claim element [7e] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Raitio-Fahlman combination renders obvious [7e] for reasons discussed for [1e]-[1f].

- 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.”**

136. Claim element [7f] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Raitio-Fahlman combination renders obvious [7f] for reasons discussed for [1e]-[1f].

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

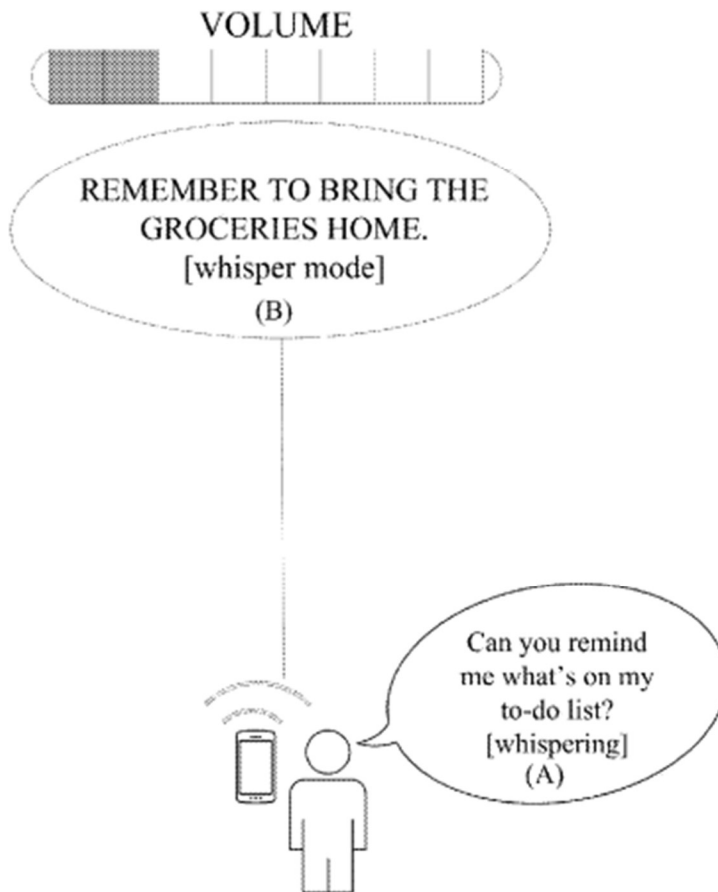
1. Independent Claim 1

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

137. Ocampo discloses [1preamble] because it discloses a “user device” for “convert[ing] the content in [a] received message to an audio signal and control[ing] the output of [an] audio signal to be at a volume proportional to [a] determined proximity indicator.” Ocampo, [0024]. 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;”

138. 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.”).

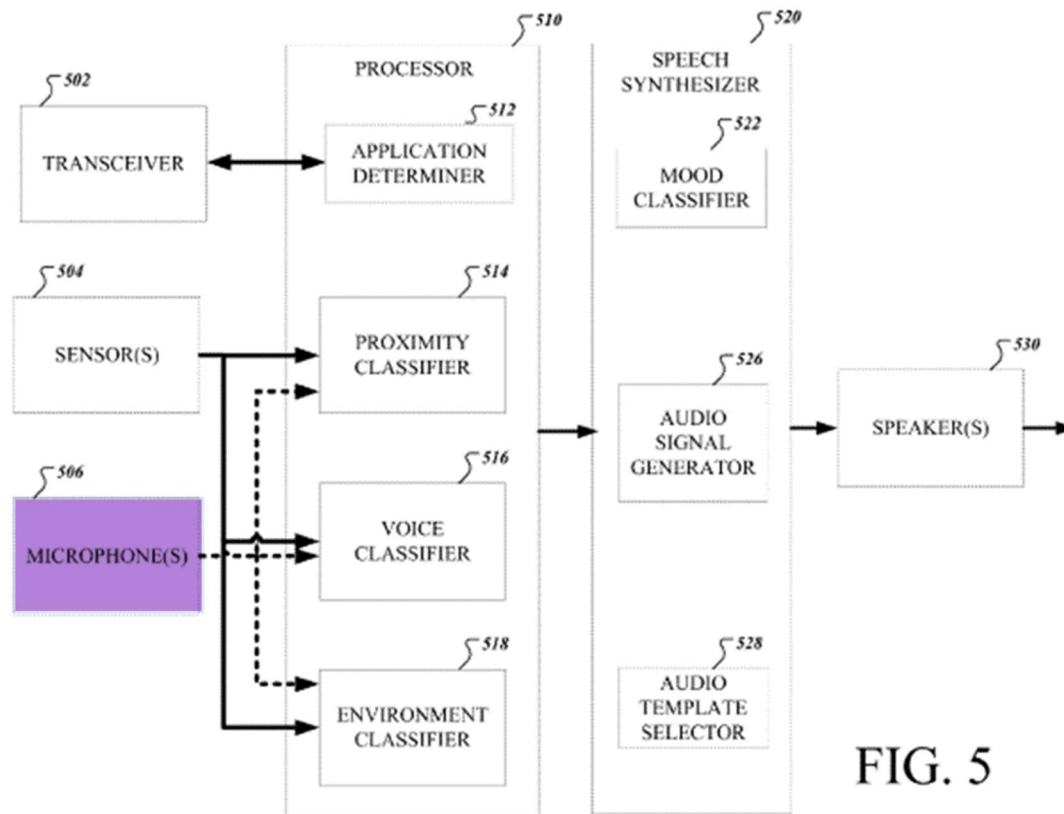


FIG. 5

Ocampo, FIG. 5 (annotated).

139. This is consistent with 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;”

140. 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.

141. 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.*”).

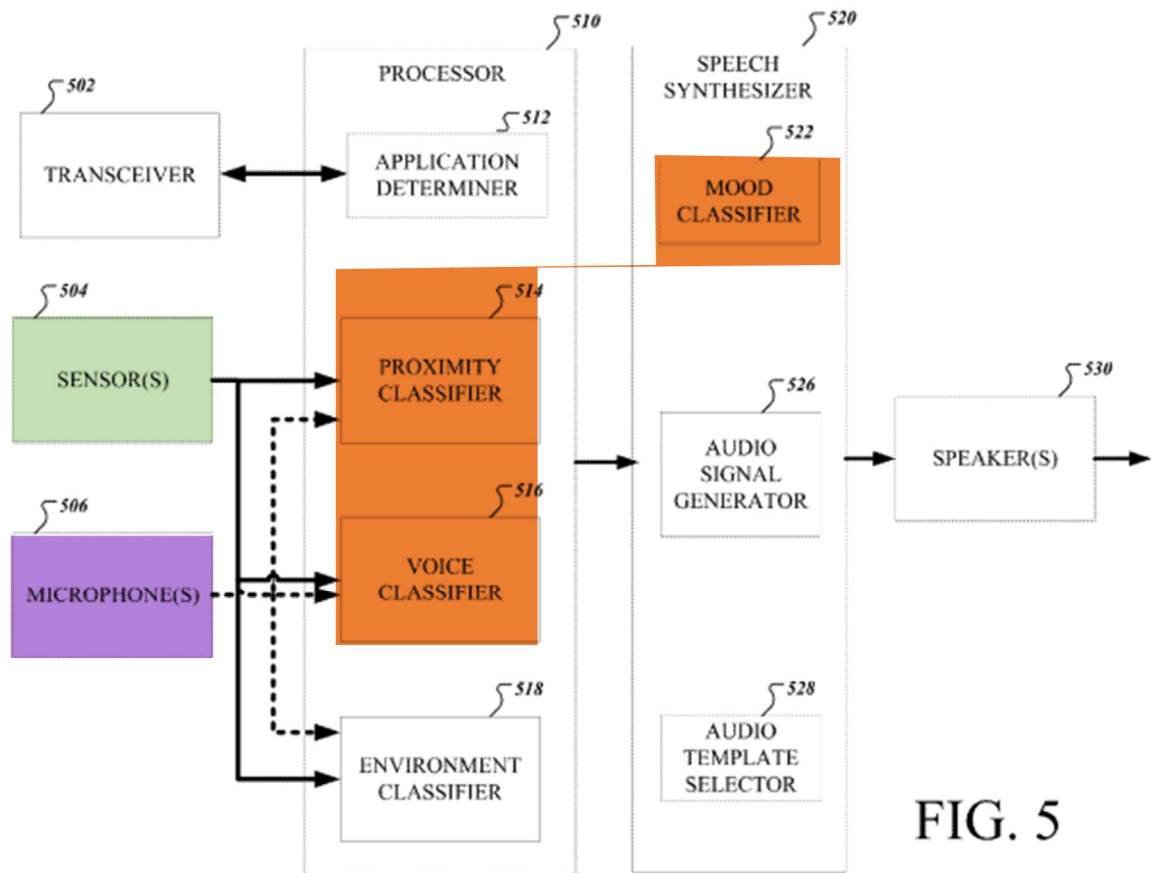


FIG. 5

Ocampo, FIG. 5 (annotated).

142. 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”).

143. 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].

144. Thus, it is my opinion that 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).

- 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;”**

145. 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].

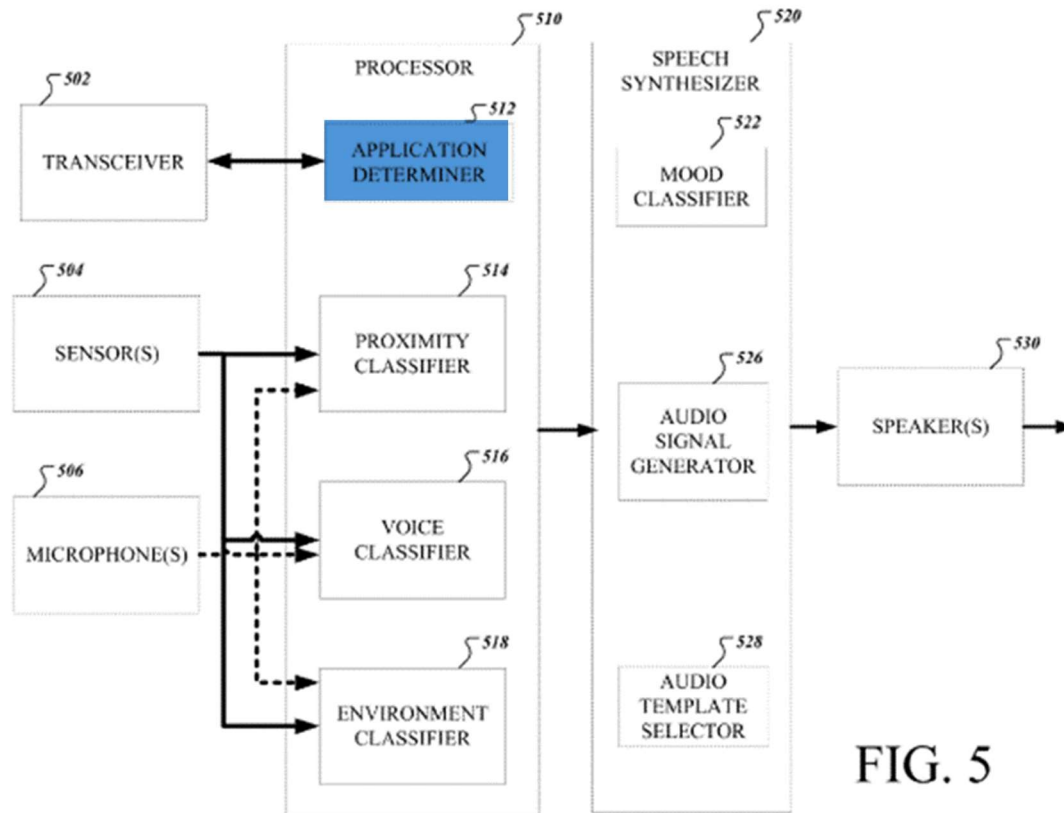
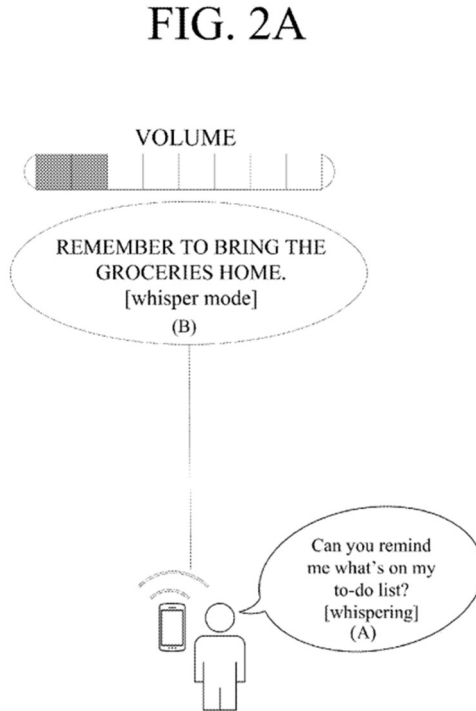


FIG. 5

146. 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]. 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].

147. 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]. 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.

148. Accordingly, it is my opinion that 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 I further explain in [1d], below.

Ocampo, [0036]-[0037], [0051], FIG. 2A.

149. 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"**

150. 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].

151. As I explained in Ground 2, [1c], above, 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]. 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 exact source of the data "depends on various factors [,] including the type of application and type of command" and the data itself "may be retrieved in various suitable ways [,] including, for example, 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.

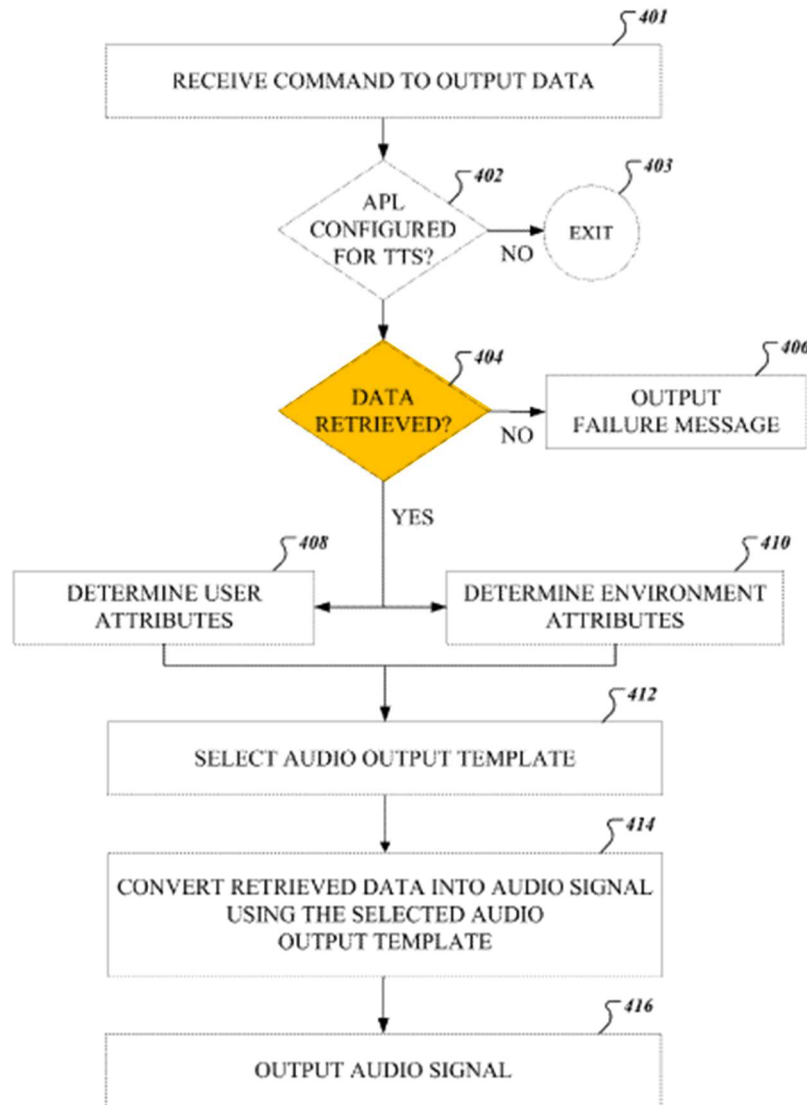


FIG. 4

Ocampo, FIG. 4 (annotated).

152. 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, it is my opinion that 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 I further explain in [1e], below. Ocampo, [0056].

- 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,”**

153. 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].

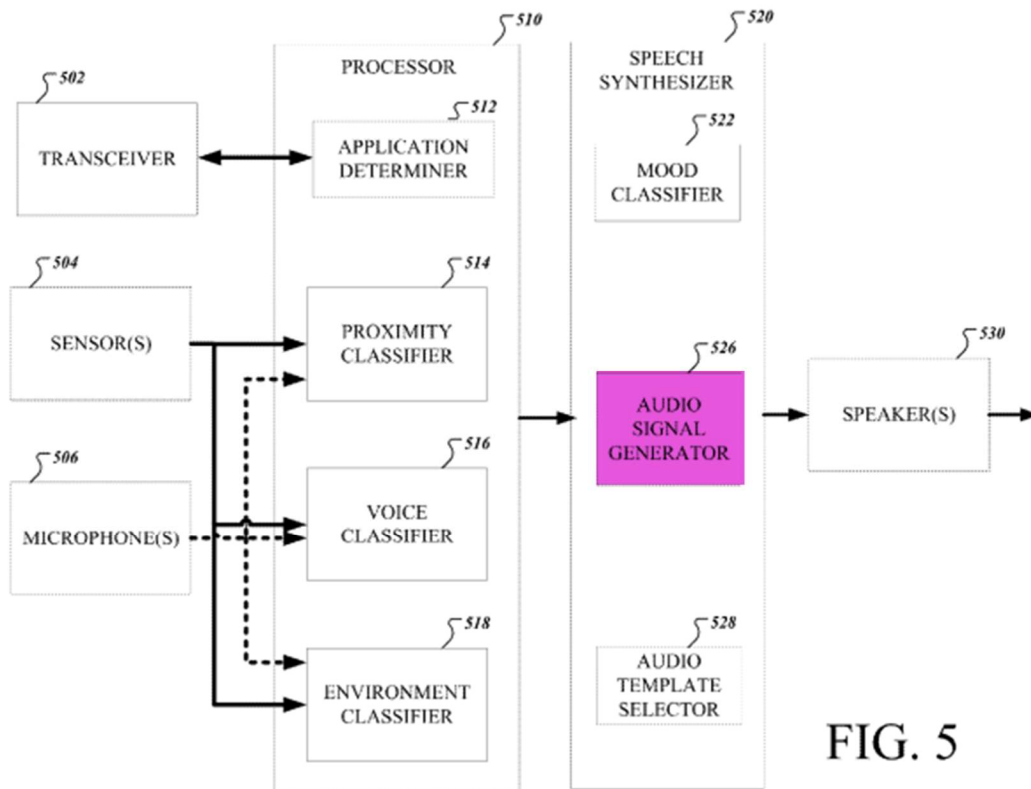


FIG. 5

Ocampo, FIG. 5 (annotated).

154. As I explain for Ground 2, [1d], above, 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.

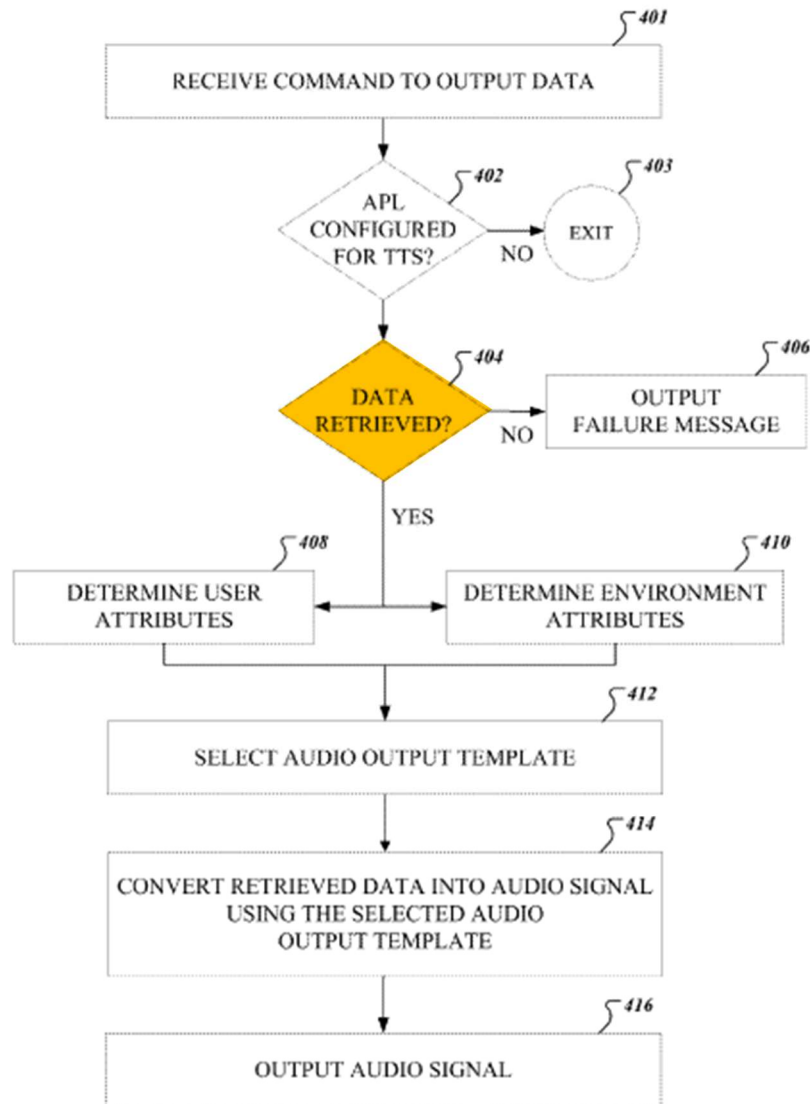


FIG. 4

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.

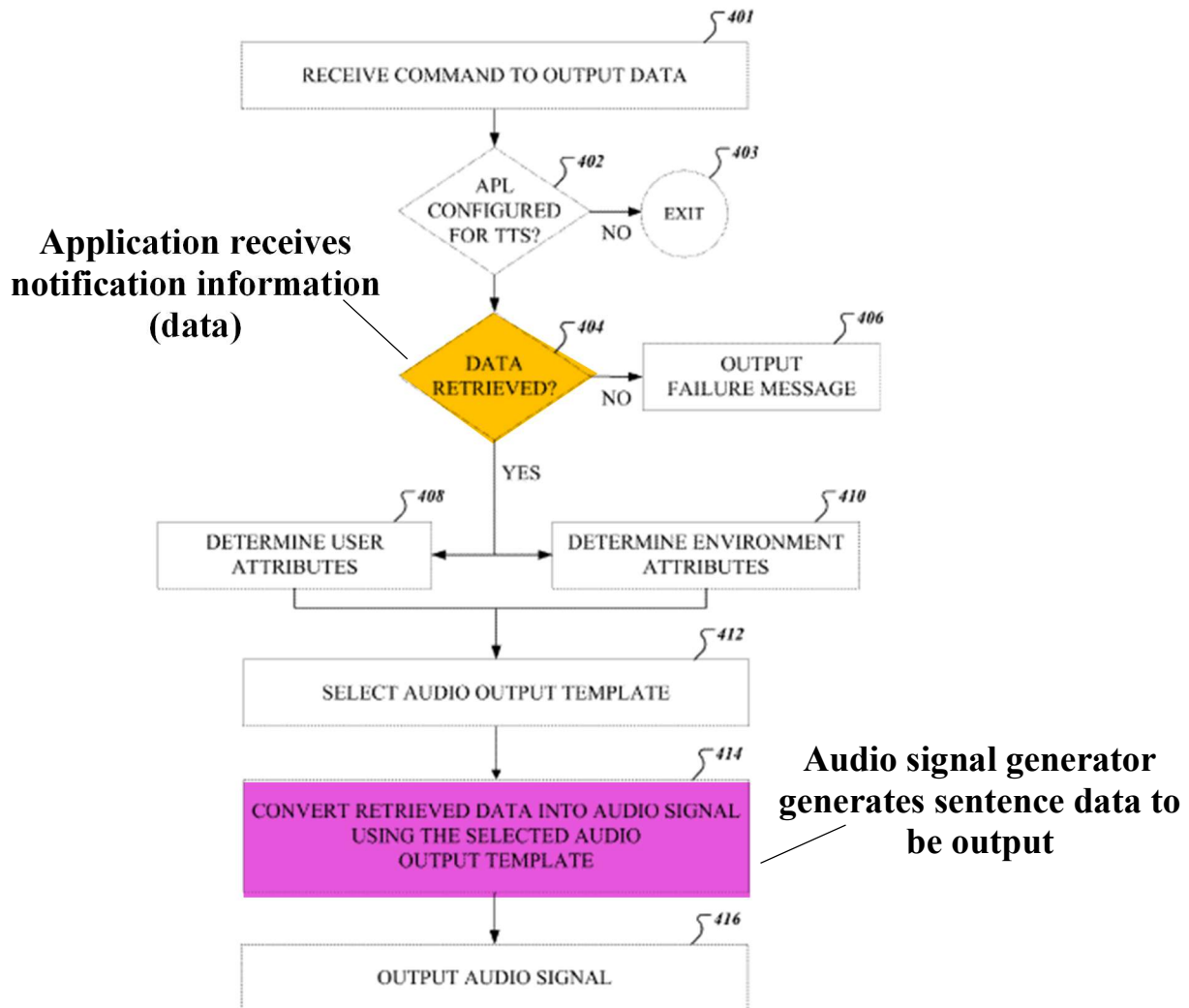


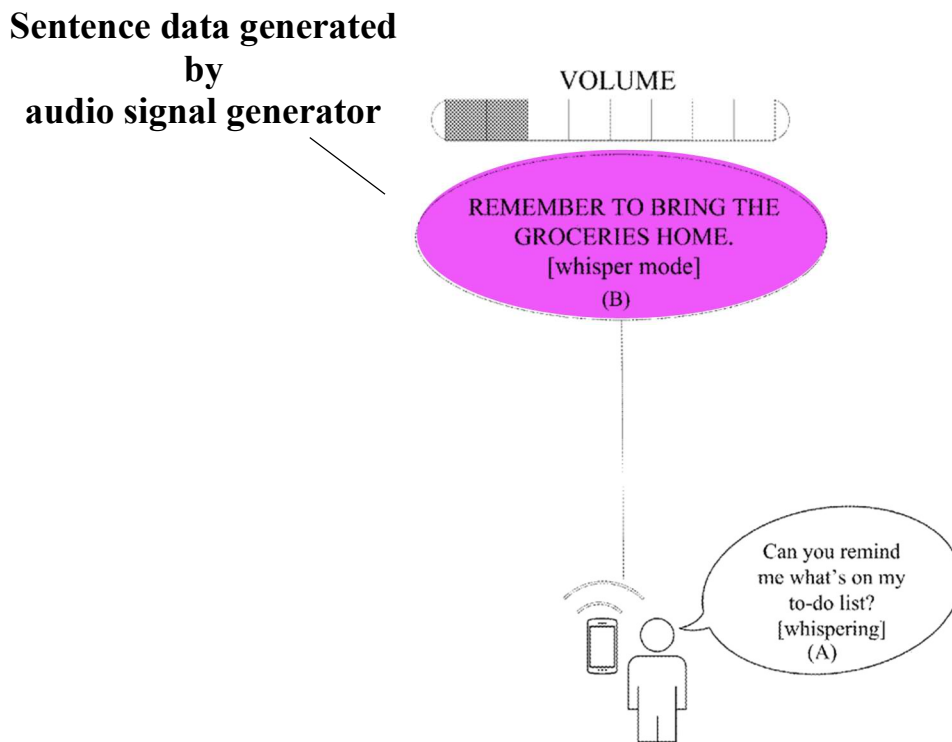
FIG. 4

Ocampo, FIG. 4 (annotated).

155. Referring back to the exemplary scenario of Fig. 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).

156. 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 I explain in further detail in [1f], below. 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.”**

157. In my opinion, Ocampo-Fahlman renders obvious [1f].

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

159. 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

160. 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. 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.

161. Thus, it is my opinion that 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.

ii. Expectation of Success

162. It is also my opinion that 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.

163. 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 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." Ocampo, [0090], [0092]. Indeed, Ocampo's "audio template selector" is a program in the user device's speech synthesizer processor. Ocampo, [0047].

164. 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.

165. 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.

2. **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.”**

166. In my opinion, Ocampo discloses, and the Ocampo-Fahlman combination renders obvious the additional limitations of claim 2.

167. First, Ocampo discloses or Ocampo-Fahlman renders obvious “*the notification information includes type information indicating a type of the content information.*” As I explain for Ground 2, [1d], above, 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 information *place*. Ocampo, [0077]. This is consistent with 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.

168. I have also been asked to consider whether Fahlman renders obvious the additional limitations of claim 2. In my opinion, it does. 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).

169. Accordingly, it is also my opinion that 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], above.

170. Second, Ratio-Fahlman renders obvious “*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.*”

171. As I explain for Ground 2 [1f], above, 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.

172. 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.

173. 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.

- 3. 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.”**

174. In my opinion, the Ocampo-Fahlman combination renders claim 3 obvious.

175. As I explain for Ground 1, claim 3, above, 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 token[s]” (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.

176. 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.

4. **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.”**

177. In my opinion, the Ocampo-Fahlman combination renders claim 4 obvious because, as I explain in Ground 1, claim 4, above, 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).

178. Fahlman’s 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.

179. 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.

5. 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.”

180. Ocampo discloses the additional limitations of claim 5 because, as I explain in Ground 2, [1b], above, its classifiers “classify audio signals received from a user and determine... *whether a user is whispering.*” Ocampo, [0035],[0037], [0065], FIG. 2A. 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.

181. It is also my opinion that Ocampo 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].

182. 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 are configured to determine that the voice belongs to a particular user (a predetermined voice).

6. Independent Claim 6

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

183. 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], above), 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].

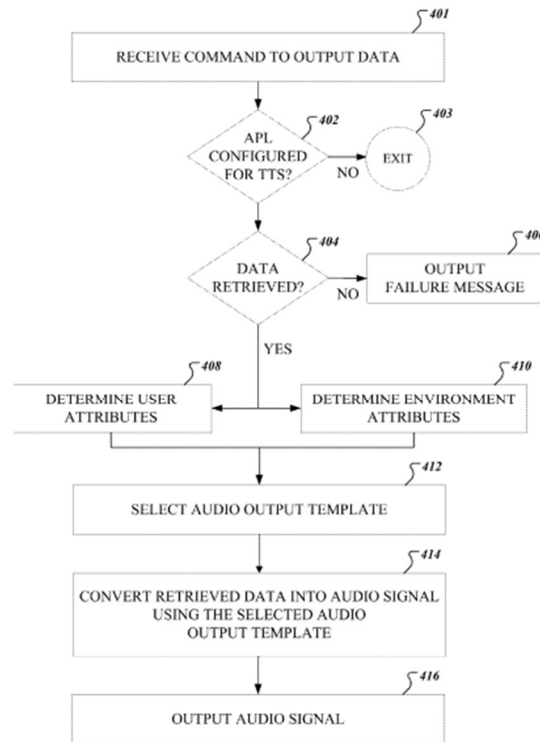


FIG. 4

Ocampo, FIG. 4.

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

184. Claim element [6a] is substantively similar to claim element [1a] above; thus, it is my opinion that Ocampo discloses [6a] for reasons discussed for Ground 2, [1a], above.

- 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;”**

185. Claim element [6b] is substantively similar to claim element [1b] above; thus, it is my opinion that Ocampo discloses [6b] for reasons discussed for Ground 2, [1b], above.

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

186. Claim element [6c] is substantively similar to claim element [1c] above; thus, it is my opinion that Ocampo discloses [6c] for reasons discussed for Ground 2, [1c], above.

- 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”**

187. Claim element [6d] is substantively similar to claim element [1d] above; thus, it is my opinion that Ocampo discloses [6d] for reasons discussed for Ground 2, [1d], above.

- 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”**

188. Claim element [6e] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Ocampo-Fahlman combination renders obvious for reasons discussed for Ground 2, [1e]-[1f], above.

- 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.”**

189. Claim element [6f] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Ocampo-Fahlman combination renders obvious [6f] for reasons discussed for Ground 2, [1e]-[1f], above.

7. Independent Claim 7

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

190. 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;”

191. Claim element [7a] is substantively similar to claim element [1a] above; thus, it is my opinion that Ocampo discloses [7a] for reasons discussed for Ground 2, [1a], above.

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;”

192. Claim element [7b] is substantively similar to claim element [1b] above; thus, it is my opinion that Ocampo discloses [7b] for reasons discussed for Ground 2, [1b], above.

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

193. Claim element [7]c is substantively similar to claim element [1c] above; thus, it is my opinion that Ocampo discloses [7c] for reasons discussed for Ground 2, [1c], above.

- 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”**

194. Claim element [7d] is substantively similar to claim element [1d] above; thus, it is my opinion that Ocampo discloses [7d] for reasons discussed for [1d].

- 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”**

195. Claim element [7e] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Ocampo-Fahlman combination renders obvious [7e] for reasons discussed for Ground 2, [1e]-[1f], above.

- 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.”**

196. Claim element [7f] is substantively similar to claim elements [1e]-[1f] above; thus, it is my opinion that the Ocampo-Fahlman combination renders obvious Ocampo discloses [7f] for reasons discussed for Ground 2, [1e]-[1f], above.

XI. CONCLUSION

197. For the reasons set forth in Section X, it is my opinion that one skilled in the art would have found claims 1-7 of the '675 Patent obvious.

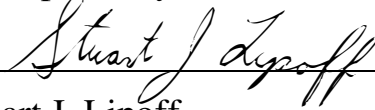
198. In signing this declaration, I understand that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I acknowledge that I may be subject to cross-examination in this case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

199. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Declaration of Stuart J. Lipoff
U.S. Patent No. 11,244,675

Date: July 13, 2025

Respectfully submitted,



Stuart J. Lipoff