

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner.

Case No. IPR2025-01123
U.S. Patent No. 11,069,337

**DECLARATION OF MR. STUART LIPOFF
UNDER 37 C.F.R. § 1.68 IN SUPPORT OF PETITION FOR
INTER PARTES REVIEW**

TABLE OF CONTENTS

I.	Introduction.....	8
II.	Qualifications and Professional Experience.....	10
III.	Relevant Legal Standards.....	17
IV.	Level Of Ordinary Skill.....	18
V.	Prior Art.....	18
VI.	Anticipation.....	19
VII.	Obviousness.....	19
VIII.	'337 Patent.....	24
	A. Overview.....	24
	B. Prosecution History.....	28
IX.	Level of Ordinary Skill in the Art.....	29
X.	Claim Construction.....	29
XI.	Claims 1-5 are Unpatentable.....	30
	A. Ground 1: Ocampo-Yi Renders Claims 1-5 Obvious.....	31
	1. Ocampo (Ex. 1005) Summary.....	31
	2. Yi (Ex. 1006) Summary.....	32
	3. Reasons to Combine Ocampo and Yi.....	35
	4. Independent Claim 1.....	38
	a) [1pre] A voice-content control device, comprising:.....	38
	b) [1a] a proximity sensor configured to calculate a distance between a user and the voice-content control device;.....	39

- c) [1b] a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device;.....42
- d) [1c] a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user;46
- e) [1d] a voice-content generating unit configured to generate, based on content of the processing executed by the process executing unit, output sentence that is text data for a voice to be output to the user; and.....48
- f) [1e] an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence therein, wherein.....51
- g) [1f] the voice-content generating unit is further configured to52
 - i. [1f.i] generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice, and..... 52
 - ii. [1f.ii] generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice, wherein..... 55
- h) [1g] the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence therein differs from the sound

volume of the voice data obtained by
converting the second output sentence
thereinto.61

5. Claim 2.....62

a) [2a]: The voice-content control device according
to claim 1, wherein the process executing
unit comprises: an intention analyzing unit
configured to extract intention information
indicating an intention of the user based on
the voice acquired by the voice acquiring
unit; and62

b) [2b]: an acquisition content information acquiring
unit configured to acquire acquisition
content information which is notified to the
user based on the extracted intention
information, and.....63

c) [2c]: the voice-content generating unit is further
configured to generate the text data
including the acquisition content
information as the output sentence.64

6. Claim 3: The voice-content control device according to claim
1, wherein, on generating the second sentence, the voice-
content generating unit is further configured to omit a part of
information included in the voice spoken by the user.....65

7. Independent Claim 4.....67

a) [4pre] A voice-content control method,
comprising:67

b) [4a] calculating a distance between a user and a
voice-content control device;.....67

c) [4b] acquiring a voice spoken by a user;67

d)	[4c] analyzing the acquired voice to classify the acquired voice as either one of a first voice and a second voice based on the distance between the user and the voice-content control device;	68
e)	[4d] analyzing the acquired voice to execute processing intended by the user;.....	68
f)	[4e] generating, based on content of the executed processing, [an] output sentence that is text data for a voice to be output to the user; and.....	68
g)	[4f] adjusting a sound volume of voice data obtained by converting the output sentence thereinto, wherein at the generating,	68
	i. [4f.i] a first output sentence is generated as the output sentence when the acquired voice has been classified as the first voice, and.....	68
	ii. [4f.ii] a second output sentence is generated as the output sentence in which a part of information included in the first output sentence is omitted when the acquired voice has been classified as the second voice, wherein	69
h)	[4g] at adjusting the sound volume of voice data, further adjusting the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of voice data obtained by converting the second output sentence thereinto.	69
8.	Independent Claim 5.....	69

- a) [5pre] A non-transitory storage medium that stores a voice-content control program that causes a computer to execute:69
- b) [5a] calculating a distance between a user and a voice-content control device;.....70
- c) [5b] acquiring a voice spoken by a user;70
- d) [5c] analyzing the acquired voice to classify the acquired voice as either one of a first voice and a second voice based on the distance between the user and the voice-content control device;70
- e) [5d] analyzing the acquired voice to execute processing intended by the user;.....70
- f) [5e] generating, based on content of the executed processing, [an] output sentence that is text data for a voice to be output to the user; and.....70
- g) [5f] adjusting a sound volume of voice data obtained by converting the output sentence thereinto, wherein at the generating,71
 - i. [5f.i] a first output sentence is generated as the output sentence when the acquired voice has been classified as the first voice, and..... 71
 - ii. [5f.ii] a second output sentence is generated as the output sentence in which a part of information included in the first output sentence is omitted when the acquired voice has been classified as the second voice wherein 71
- h) [5g] at adjusting the sound volume of voice data, further adjusting the sound volume of voice data such that the sound volume of voice data obtained by converting the first output

	sentence thereinto differs from the sound volume of voice data obtained by converting the second output sentence thereinto.	71
XII.	Conclusion	71

I, Stuart Lipoff, do hereby declare as follows:

I. INTRODUCTION

1. I make this declaration based upon my own personal knowledge and, if called upon to testify, would testify competently to the matters stated herein.

2. I have been asked by Google LLC (“Google”) to provide my expert opinion in connection with the *Inter Partes* Review of U.S. Patent No. 11,069,337 (“the ’337 patent”) (Ex. 1001) to Naganuma concerning whether the ’337 patent is unpatentable over certain prior art. This declaration is a statement of my opinions on issues relating to the patentability of claims 1-5 (“the Challenged Claims”) of the ’334 patent. As I explain more fully below, it is my opinion that all of the Challenged Claims would have been obvious to a person of ordinary skill in the art (“POSITA”) at the time of the alleged invention.

3. I am being compensated for my work in this matter at my standard hourly rate. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this investigation. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

Declaration of Mr. Stuart Lipoff
Inter Partes Review of U.S. Patent No. 11,069,337

4. I have reviewed and considered the following documents in connection with my analysis of the '337 patent:

Exhibit	Description
Ex. 1001	U.S. Patent No. 11,069,337 B2 to Naganuma (“the '337 patent”)
Ex. 1002	File History of the '337 Patent
Ex. 1005	U.S. Patent Application Publication No. US 2018/0122361 A1 to Silveira Ocampo (“Ocampo”)
Ex. 1006	U.S. Patent Application Publication No. US 2014/0303971A1 to Yi et al. (“Yi”)
Ex. 1007	U.S. Patent Application Publication No. US 2014/0025383 A1 to Dai et al. (“Dai”)
Ex. 1008	<i>SoundClear Technologies LLC v. Google LLC</i> , Case No. 1:24-cv-01281, Complaint for Patent Infringement (E.D. Va. July 25, 2024)
Ex. 1009	<i>SoundClear Technologies LLC v. Google LLC</i> , Case No. 3:24-cv-00540, Order Granting Motion to Stay (E.D. Va. Apr. 18, 2025)
Ex. 1010	U.S. Patent No. 8,468,244 B2 to Redlich et al.
Ex. 1011	U.S. Patent No. 10,089,287 B2 to Rebstock et al.
Ex. 1012	U.S. Patent No. 10,552,617 B2 to Han et al.
Ex. 1013	U.S. Patent No. 10,853,570 B2 to Matichuk et al.
Ex. 1014	U.S. Patent Application Publication No. 2017/0358301 A1 to Raitio et al. (“Raitio”)

5. In forming the opinions expressed below, I have considered:
- a) the documents listed above;
 - b) the relevant legal standards, including the standard for

obviousness, and any additional authoritative documents as cited in the body of this declaration; and

- c) my own knowledge and experience based upon my work in the field of software systems as described below, as well as the following materials.

6. 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.

7. Unless otherwise noted, all emphasis in any quoted material has been added.

II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

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

9. 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.

10. 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”).

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

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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 that 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.

19. 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.

20. 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.

21. 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.

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

23. 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.

24. 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.

25. Over the course of 15 years starting in the early 1980s, 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.

26. 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.

27. I understand my *curriculum vitae* will become provided as Exhibit 1004.

III. RELEVANT LEGAL STANDARDS

28. 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.

29. 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.

30. 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).

31. I have been asked to provide my opinions as to whether the cited prior art discloses or renders obvious claims 1-15 of the '374 Patent from the perspective of a POSITA at the time of the earliest claimed priority date of February 20, 2012, as described in more detail below.

32. 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.”

33. 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.

IV. LEVEL OF ORDINARY SKILL

34. 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.

35. 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.

V. PRIOR ART

36. 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.

VI. ANTICIPATION

37. 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.

VII. OBVIOUSNESS

38. 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.

39. 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).

40. 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 February 20, 2012. My opinions would not change if I assumed another, e.g., later, date of invention.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

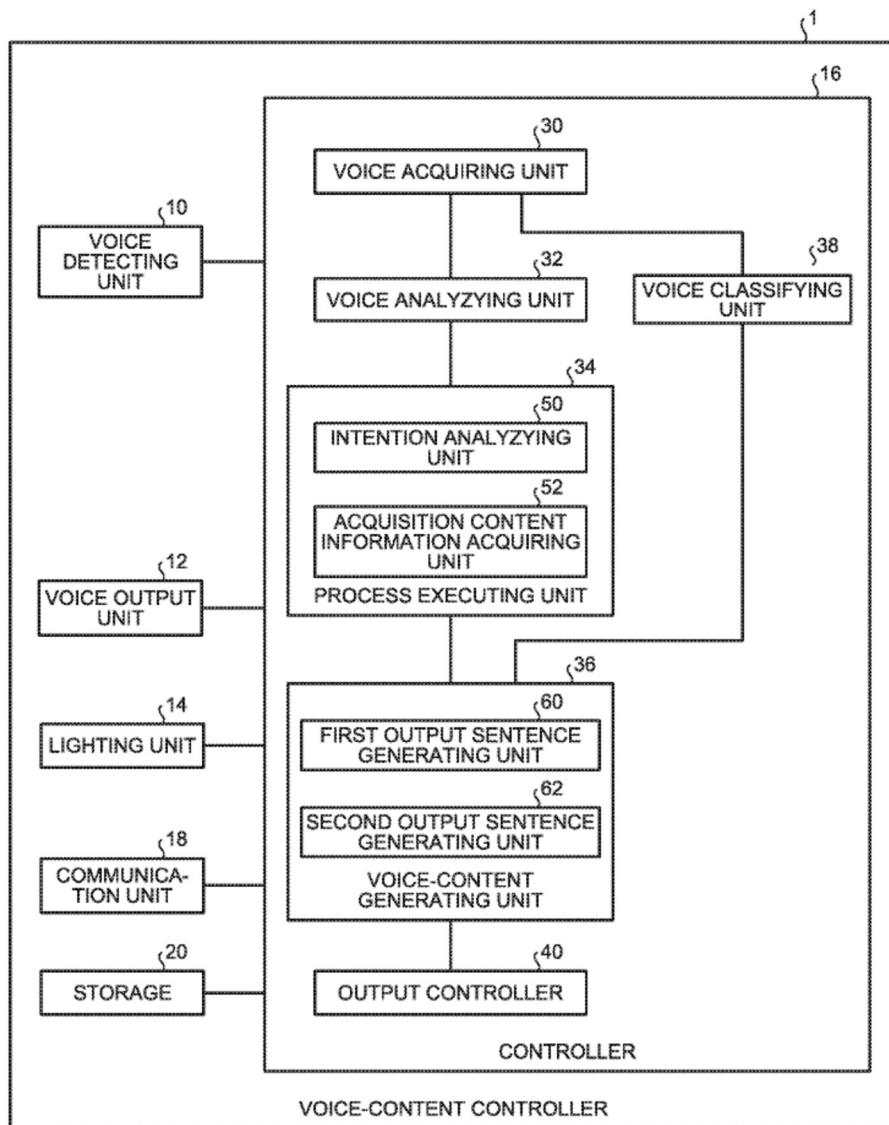
VIII. '337 PATENT

A. Overview

49. The '337 patent, filed on March 4, 2019, and issued on July 20, 2021, claims priority to Japanese Patent Application No. JP2018-039754 filed March 6, 2018. Ex. 1001, cover, (22), (30), (45). The '337 patent relates to a “voice-content control device” that generates responses to spoken user commands based on an analysis of the user’s voice. Ex. 1001, 1:48-64, Abstract. Figure 2 below illustrates a

voice-content control device for receiving the spoken user command, processing it to determine voice attributes and identify responsive information, and generating an audible response to the command. Ex. 1001, Abstract, FIG. 2.

FIG.2



Ex. 1001, FIG. 2.

50. The device includes a “voice detecting unit 10” configured to receive

a user's spoken "wishes" and a "controller 16" comprised of various units that allow the device to: extract "intention information" that "indicates what kind of processing is intended by the user H to be performed"; classify the voice as a "first voice" or "second voice," e.g., a whisper or non-whisper, based on the user's distance; and generate an output based on the voice's classification and the user's "wish[]." Ex. 1001, 3:13-59, 4:15-5:15, 7:51-8:34, 8:60-10:18, FIGS. 5-6.

51. For example, an "intention analyzing unit 50" may "extract[]" the intention information" from text data corresponding to the user's voice command using "natural language processing" (e.g., Ex. 1001, 4:31-33); a "voice classifying unit 38" may classify a user's voice "V1" as a "first voice V1A when the voice V1 is determined not to be a whisper" and "second voice V1B when the voice V1 is determined to be a whisper." (Ex. 1001, 7:51-63); and an "acquisition content information acquiring unit 52" to execute the "processing required by the user" (Ex. 1001, 6:24-57, FIG. 2).

52. The '337 patent also describes a "voice-content generating unit 36" for generating an output sentence that responds to the user command using a "first output sentence" or a "second output sentence," the latter omitting information as compared to the first output sentence depending on whether the user's voice is classified as voice V1A (non-whisper) or V1B (whisper). Ex. 1001, 7:51-63. To determine whether the user whispered a command, "the distance [between the user

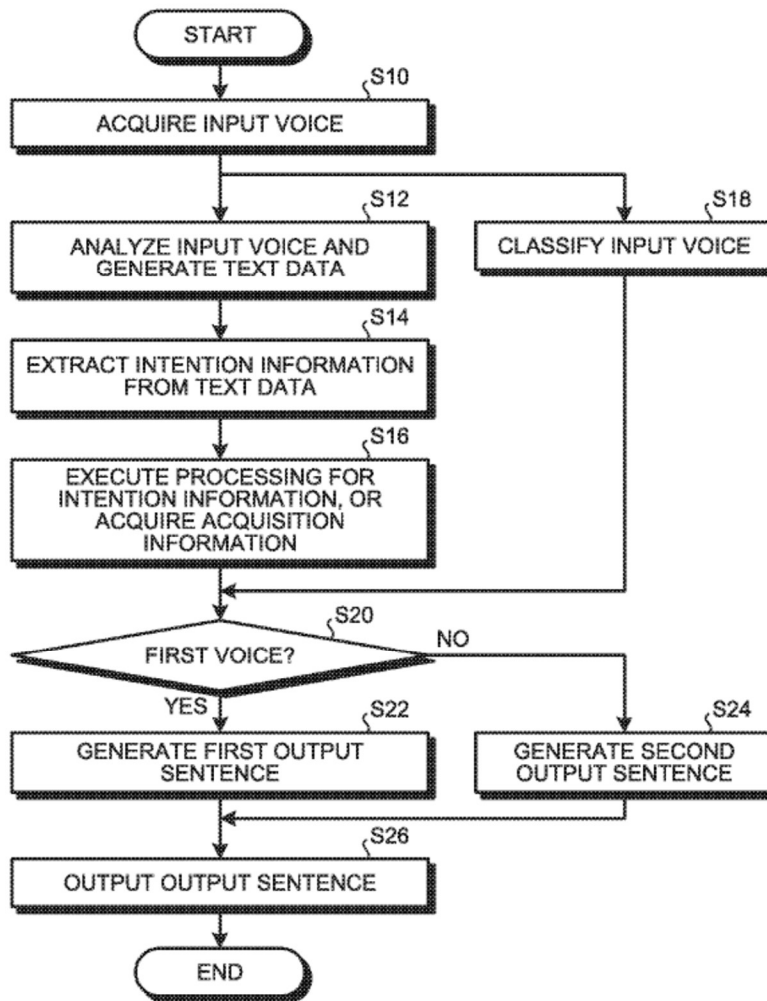
and device] can be used as a feature value to perform the classification to the first voice V1A and the second voice V1B.” Ex. 1001, 8:20-26.

53. “By detecting the whispered voice to generate the second output sentence, the voice-content control device 1 can determine whether the voice-content control device 1 is in a state of influencing any person other than the user H adequately, and suppress the influence adequately.” Ex. 1001, 17:11-18.

Therefore, the “output controller 40” converts output sentences into voice with “different sound volumes” depending on the output sentence. Ex. 1001, 13:50-14:2.

54. Figure 5 illustrates a process for generating a first or second output sentence based on voice classification and user intent. Ex. 1001, 14:5-60, FIG. 5.

FIG.5



Ex. 1001, FIG. 5.

B. Prosecution History

55. The '337 patent began as U.S. Application No. 16/290,983, filed March 4, 2019, with a claim of priority from JP Application No. JP2018-039754 filed March 6, 2018. See Ex. 1001, cover, (21)-(22), (30); Ex. 1002, 330-336. The Office initially rejected then-pending claims 1 and 3-6 as anticipated, and claim 2

as obvious, in view of the prior art of record. See Ex. 1002, 195-204. In response, the Applicant amended the independent claims to include an output controller that alters the volume between different spoken response sentences, and argued that the prior art lacked the newly claimed subject matter. Ex. 1002, 183, 186-191. The Office thereafter issued a Final Office Action rejecting claims 1-6 as obvious over additional prior art. Ex. 1002, 162-176. The Applicant then amended the claims to include a proximity sensor and to require that voice classification (i.e., as first voice or second voice) occur based on the distance between the user and device. Ex. 1002, 153-159. The Office allowed the claims as amended. Ex. 1002, 96-102.

IX. LEVEL OF ORDINARY SKILL IN THE ART

56. In my opinion, a person of ordinary skill in the art (“POSITA”) in the field of the ’337 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.

X. CLAIM CONSTRUCTION

57. It is my understanding that in order to properly evaluate the ’374 patent, the terms of the claims must first be interpreted. It is my understanding that for the purposes of this inter partes review, the claims are to be construed under the so-called *Phillips* standard, under which claim terms are given their ordinary and

customary meaning as would have been understood by one of ordinary skill in the art in light of the specification and prosecution history, unless the inventor has set forth a special meaning for a term.

XI. CLAIMS 1-5 ARE UNPATENTABLE

58. I have been asked to provide my opinion as to whether the Challenged Claims of the '337 patent would have been obvious in view of the prior art. The discussion below provides a detailed analysis of how the prior art references identified below teach the limitations of the Challenged Claims of the '337 patent.

59. As part of my analysis, I have considered the scope and content of the prior art and any differences between the alleged invention and the prior art. I describe in detail below the scope and content of the prior art, as well as any differences between the alleged invention and the prior art, on an element-by-element basis for each Challenged Claims of the '337 patent.

60. As described in detail below, the alleged invention of the Challenged Claims would have been obvious in view of the teachings of the identified prior art references as well as the knowledge of a POSITA.

Grounds	Claims	Basis	Prior Art
1	1-5	§ 103	Ocampo and Yi

A. Ground 1: Ocampo-Yi Renders Claims 1-5 Obvious

1. Ocampo (Ex. 1005) Summary

61. I understand that because Ocampo published on May 3, 2018, it qualifies as prior art to the '337 patent under 35 U.S.C. § 102(a)(1). Ocampo, cover, (22), (43). Ocampo discloses a user device having dynamic text-to-speech (TTS) provisioning that automatically controls voice output in response to user commands. Ocampo, Abstract, [0003]. Ocampo classifies the voice according to user attributes, including the voice characteristics indicating temperament (e.g., whispering) and proximity to the microphone, and selects audio output based on the classification (e.g., whispering classification) to adjust the volume of voiced output responding to the user. Ocampo, [0060]-[0065], [0087], [0088], FIG. 5.

62. User attributes include various voice features such as pitch, tone, frequency, and amplitude, which are extracted from the user's voice signals. Ocampo, [0005]-[0006]. Ocampo's user device also accounts for the user's proximity to the device, determined through audio signals received by multiple microphones and sensor data. Ocampo, [0060]-[0061]. The device classifies the user's likely mood based on these user attributes, allowing it to tailor the audio output accordingly. Ocampo, [0065]. For example, Ocampo's device identifies the user has whispered close to the device by detecting low volume and pitch in the user's voice to adjust the responsive audio output accordingly. Ocampo, [0035],

FIGS. 2A, 4.

63. For example, the user device dynamic adjusts the volume and tone for the audio output responding to a user command based on the user's proximity.

Ocampo, [0009]-[0011], [0023]-[0024]. When the user is close to the device and in a quiet environment, for example, the system outputs the audio responding at a lower volume. Ocampo, [0027], [0038], [0075], FIG. 2A. Conversely, if the user is farther away or in a noisy environment, the volume is increased. Ocampo, FIGS. 1B, 2B.

64. Ocampo is analogous art to the '337 patent because it is in the same field of endeavor as the '337 patent: speech processing. *Compare* Ocampo, Abstract, *with* Ex. 1001, Abstract. Furthermore, Ocampo is directed to the same problem as the '337 patent: analyzing voice input commands to generate responses. *Compare* Ocampo, Abstract, *with* Ex. 1001, Abstract.

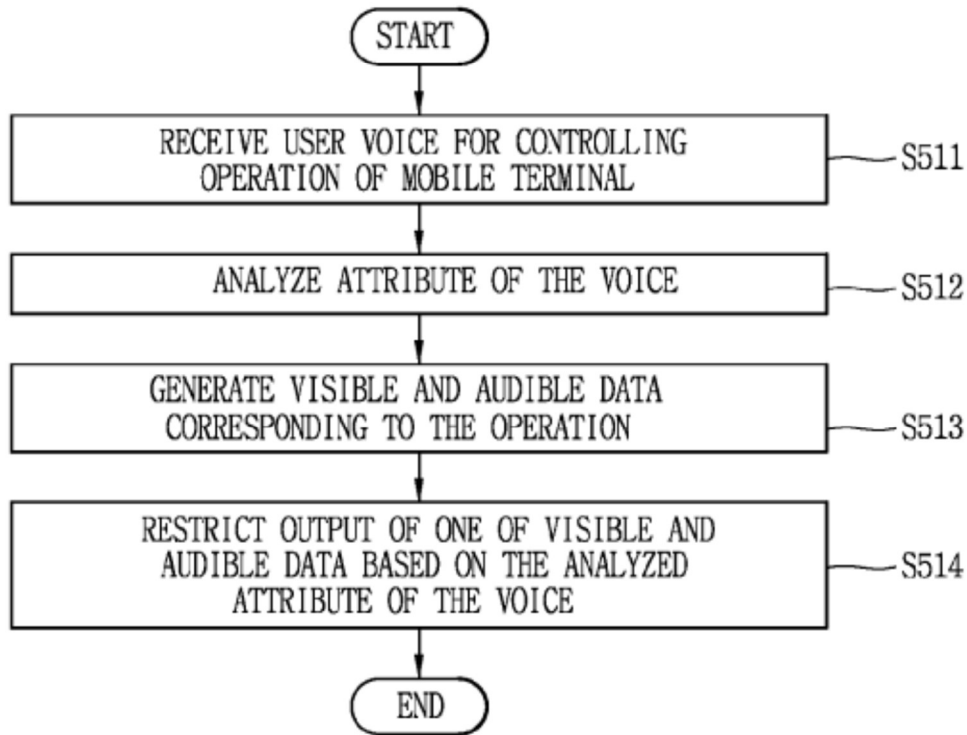
2. Yi (Ex. 1006) Summary

65. I understand that because Yi published on October 9, 2014, it also qualifies as prior art to the '337 patent under 35 U.S.C. § 102(a)(1). Yi, cover, (22), (43). Yi discloses a mobile terminal that restricts the output and volume of its vocalized responses to user commands when “the voice is sorted as . . . whispering” and “the degree of proximity is smaller than a preset range,” to account for user privacy. Yi, Abstract, [0186]-[0188].

66. Yi explains that the mobile terminal includes “an analyzing unit which analyzes an attribute of a voice input through the microphone 122.” Yi, [0167]. Yi’s mobile terminal analyzes the voice input based on various criteria to identify attributes of the voice, including whisper status and proximity to the microphone. Yi, [0164], [0167], [0169]-[0171], [0182], [0184], [0253], FIGS. 4 (S512, S513), 5B (S512c). Yi also explains that “a proximity sensor 141 may be arranged at an inner region of the mobile terminal” using various technologies to sense the presence, absence, or approach of the user with respect to the mobile terminal. Yi, [0085]-[0091].

67. Yi also discloses “analyz[ing] the meaning of the user’s voice 610, and control[ling] the mobile terminal to execute an operation according to the voice 610.” Yi, [0164]. This includes “generat[ing] visible and audible data based on the voice 610.” Yi, [0164].

FIG. 4



Yi, FIG. 4.

68. For example, where a user asks, “What’s the weather like today?” Yi’s controller 180 “may execute a function for providing information related to today’s weather” and “generate data to be provided to the user among the [responsive] information.” Yi, [0164], FIG. 5A. Moreover, the mobile device “may control the output unit to restrict (limit) an output” of visible and audible data based on the user’s voice. Yi, [0167], FIG. 4 (S514). This includes omitting

information from the voiced response to the user's command and controlling its volume. Yi, [0186]-[0188], [0221], FIG. 5A.

69. Yi is analogous art to the '337 patent because it is in the same field of endeavor as the '337 patent: speech processing. *Compare* Yi, Abstract, *with* Ex. 1001, Abstract. Furthermore, Yi is directed to the same problem as the '337 patent: analyzing voice input commands to generate responses. *Compare* Yi, Abstract, *with* Ex. 1001, Abstract.

3. Reasons to Combine Ocampo and Yi

70. In my opinion, both Ocampo and Yi are analogous art to the '337 patent because they are in the same field of endeavor as the '337 patent: speech processing. *Compare* Ocampo, Abstract, and Yi, Abstract, *with* Ex. 1001, Abstract. Furthermore, Ocampo and Yi are both directed to the same problem as the '337 patent: analyzing voice input commands to generate responses. *Compare* Ocampo Abstract, *with* Ex. 1001, Abstract.

71. In my opinion, a POSITA would have been motivated to combine Ocampo and Yi and would have had a reasonable expectation of success in doing so. Both Ocampo and Yi relate to user devices for audibly responding to user voice commands. *See* Ocampo, Abstract, [0001]-[0003]; Yi, Abstract, [0010]. Ocampo describes using text-to-speech (TTS) functionality and controlling the audio output of information responsive to a user's voice command (Ocampo, [0002]-[0003]),

which a POSITA would have understood to be functionally equivalent to revising “audible data” generated in response to the user’s voiced command in Yi. Yi, [0008]. Both references account for user privacy when generating audible responses to the user. Ocampo, [0071]; Yi, [0170]-[0174], [0180], [0188]. Moreover, Ocampo describes converting responsive text information into an audio signal at a volume dependent on the user’s distance from the device (Ocampo, [0011], [0023]-[0024]), and Yi similarly describes adjusting the volume level of sound signals converted from responsive text based on the user’s proximity to the microphone (Yi, [0166]).

72. Given these similarities, it is my opinion a POSITA in possession of Ocampo would have been motivated to find references like Yi to investigate other techniques for discreetly voicing a response to a user command. For example, Ocampo explains that privacy policies may be implemented “to maintain user privacy and not output information to third parties” (e.g., Ocampo, [0071]), and a POSITA would have understood that privacy considerations other than volume—such as omitting extraneous information from the response—would further facilitate the discreet voicing of audible responses to user commands.

73. Managing sensitive information to limit exposure to parties not intended to receive it was an abundantly well-known concept to those of skill of the art in 2018. Indeed, this is an underlying concept behind redacting sensitive

information based on user identification and classification long known to the industry. *See, e.g.*, Ex. 1010 (U.S. Patent No. 8,468,244 - Redlich); Ex. 1011 (U.S. Patent No. 10,089,287 – Rebstock); Ex. 1012 (U.S. Patent No. 10,552,617 - Han); Ex. 1013 (U.S. Patent No. 10,853,570 - Matichuk). Those of skill also understood the desirability of omitting and replacing words from voiced responses when, for example, the user whispered a command near the device for security/privacy reasons. *See, e.g.*, Ex. 1014 (U.S. Patent Application Publication No. 2017/0358301 - Raitio). 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”). Moreover, it was well-known to one of skill in the art in 2018 to add or remove modal particles from voiced responses for brevity, tone, and/or emphasis. Ex. 1007 (U.S. Patent Application Publication No. 2014/0025383 - Dai). Indeed, as Yi itself teaches, restricting the output of audible data “prevent[s] data from being transferred even to another user” in quiet settings where the user has whispered a command, recognizing that “a user transfers a voice in [the] form of whispering when the user talks to a person close to him/her.” Yi, [0170], [0180].

74. It is therefore my opinion that a POSITA would have been motivated to incorporate Yi’s teachings into Ocampo to provide additional privacy and

extend Ocampo's functionality in discreetly responding to a user command.

Ocampo, [0071]. For example, a POSITA would have appreciated the benefit of omitting optional information from a response into Ocampo to extend its privacy options. Ocampo, [0071]. A POSITA would have found it obvious and straightforward to restrict the text-based responses voiced to users to accommodate these privacy techniques. Ocampo, [0071].

75. It is also my opinion that a POSITA would have had a reasonable expectation of success in combining Ocampo and Yi. Both references use ubiquitous speech synthesis techniques (e.g., text-to-speech (TTS)) to achieve predictable results (e.g., voicing the text as modified for response.). *See, e.g.*, Ocampo, [0001]-[0003]; Yi, [0166], [0235]. Both references also adjust the volume of responses based on distance to assist with privacy concerns. Therefore, in my opinion, a POSITA would have found it obvious to combine Ocampo and Yi, as doing so would merely combine existing elements (e.g., lowering volume and restricting the response based on user proximity) with known methods (e.g., proximity sensing, TTS, etc.) to yield predictable results (e.g., maintaining user privacy by limiting information exposed to third parties).

4. Independent Claim 1

a) [1pre] A voice-content control device, comprising:

76. Ocampo discloses the preamble because it describes a user device that

automatically controls responses voiced to the user. Ocampo, [0003], [0024], [0028]. Ocampo discloses a voice-content control device that automatically controls and modifies audio responding to user commands based on multiple factors, including the user's proximity and other voice attributes (i.e., whispered tone). Ocampo, Abstract, [0003]. For example, "[i]n response to receiving a command to provide information to a user, a device retrieves information and determines user and environment attributes including: (i) a distance between the device and the user when the user uttered the query; and (ii) voice features of the user." Ocampo, Abstract. Moreover, Ocampo explicitly states that "TTS operation executed on a user device may automatically control and modify an audio output based on multiple factors[,] including the user's voice" Ocampo, Abstract.

b) [1a] a proximity sensor configured to calculate a distance between a user and the voice-content control device;

77. Ocampo and Yi also disclose [1a]. For example, Ocampo discloses [1a] using classifiers to calculate the distance between a user and user device. Ocampo, [0011], [0060]-[0063]. Ocampo's device includes a proximity classifier 514 that provides a "proximity indicator" indicating the user's distance from the device based on an analysis of audio signal data from microphones 506 and sensor data from sensors 504. Ocampo, [0003], [0011], [0023], [0028], [0058]-[0059], [0063]-[0064], FIG. 5.

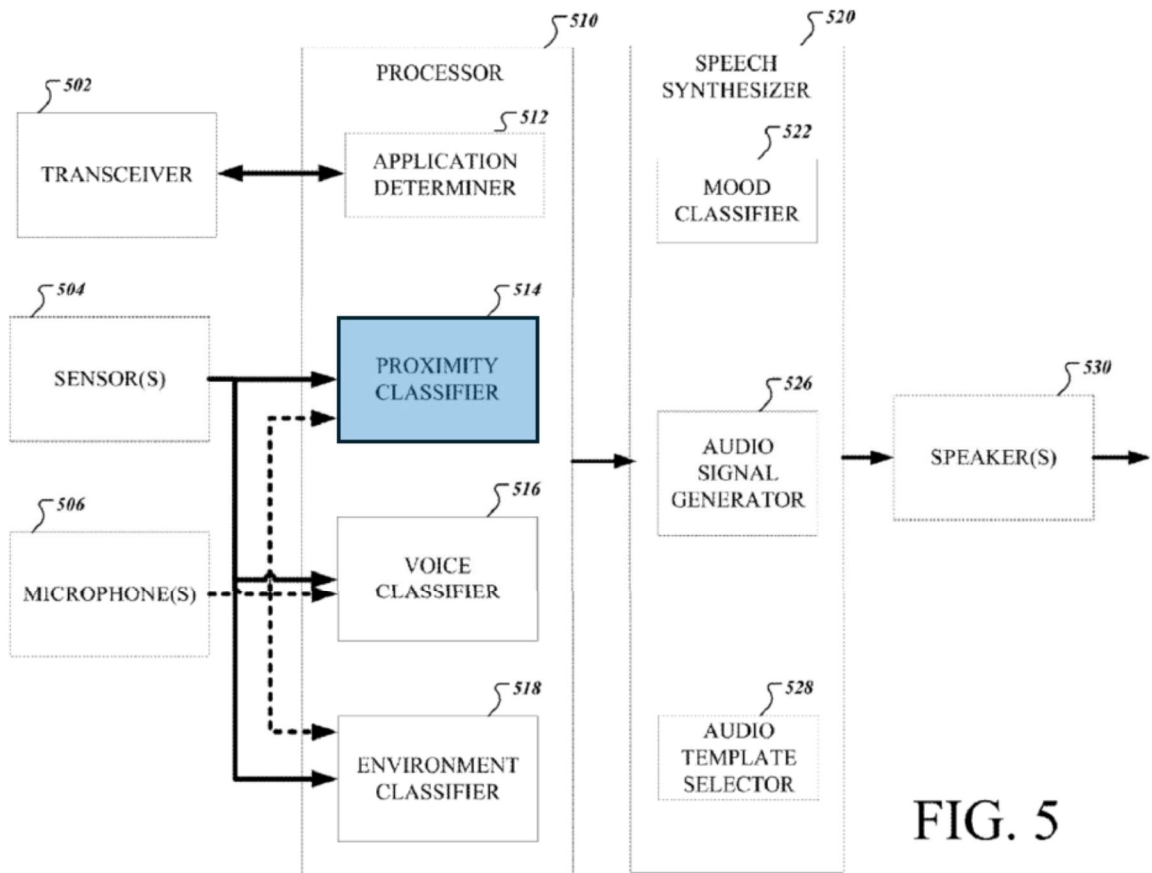


FIG. 5

Ocampo, FIG. 5 (annotated).

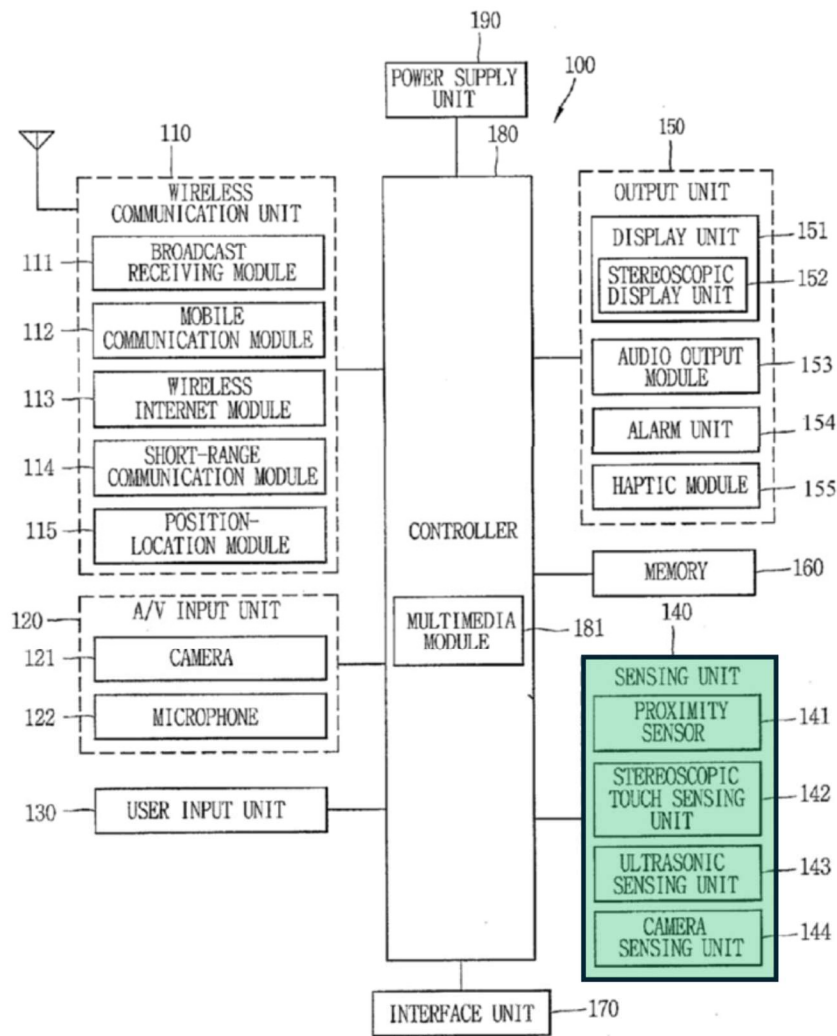
78. Ocampo also explains that its device uses the proximity indicator to, among other things, classify the user's voice for a given command by "user attributes." Ocampo, [0058]-[0065]. Ocampo explains that the proximity classifier may apply a sliding-scale analysis to determine the user's likely distance from the device based on the differences in the audio signals. Ocampo, [0062], [0063]. The proximity classifier also includes rules defining distance thresholds with the sliding scale. Ocampo, [0075], [0088].

79. Similarly, Yi discloses [1a] with its sensing unit 140 that includes a

proximity sensor 141 “configured to sense a degree of proximity between the user’s mouth and the microphone while the voice is input.” Yi, [0010], [0085],

FIG. 1.

FIG. 1



Yi, FIG. 1 (annotated).

80. Yi explains that its “proximity sensor 141 may be arranged at an inner

region of the mobile terminal” to sense the presence, absence, or approach of the user with respect to the mobile terminal using various technologies. Yi, [0085]-[0091]. But sensing unit 140 also includes a stereoscopic touch sensing unit 142, an ultrasonic sensing unit 143, and a camera sensing unit 144 for determining the mobile terminal’s proximity and orientation with respect to the user. Yi, [0090], [0092]-[0095], FIG. 1. Yi thus discloses element 1[a] too by its sensing unit 140 that provides “status measurements of various aspects of the mobile terminal” to “sense a distance between the microphone 122 and the user.” Yi, [0010], [0069], [0085], [0182], [0253]-[0255].

- c) **[1b] a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device;**

81. Ocampo and Yi disclose [1b]. Ocampo discloses [1b] with its voice classifier 516 and proximity classifier 514 for processing audio signals received at the system’s microphone(s), which become used to classify the user’s voice according to “user attributes” reflecting the user’s distance from the device among other voice features. Ocampo, [0060]-[0065], [0087], [0088], FIG. 5.

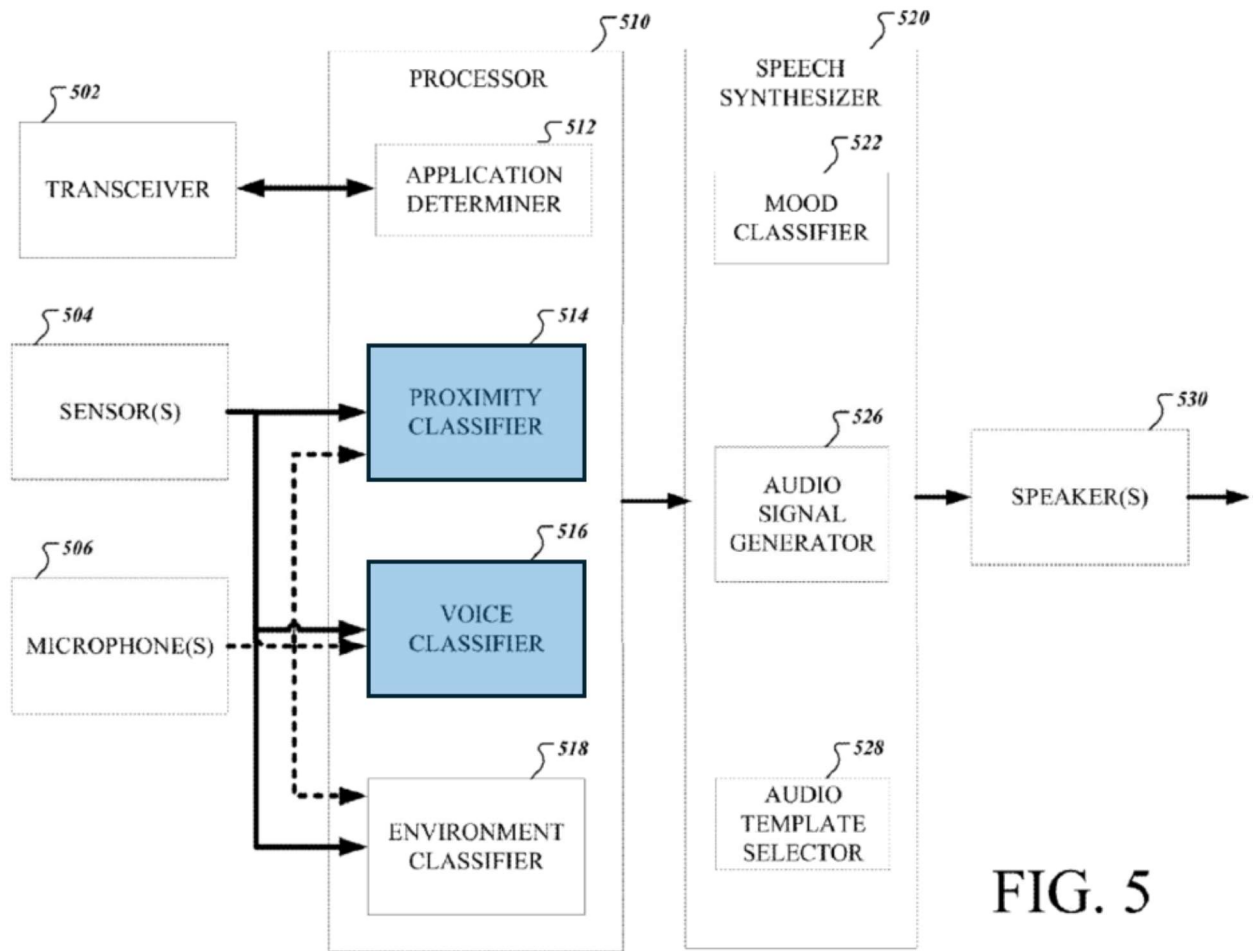


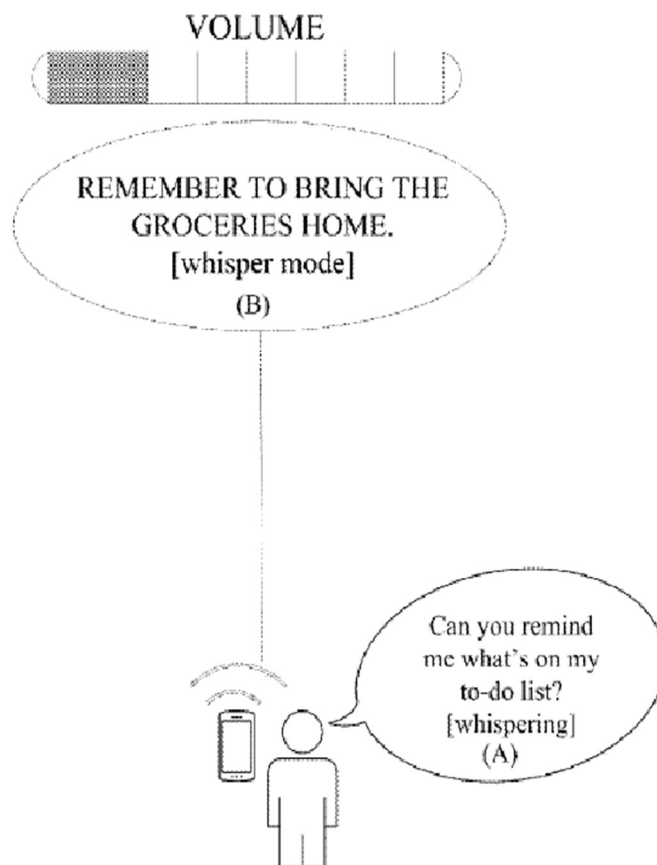
FIG. 5

Ocampo, FIG. 5 (annotated).

82. For example, Ocampo's voice and proximity classifiers extract voice features from the audio data received for a user command to determine user attributes reflecting the relative location and temperament of the user. Ocampo, [0058], [0062]-[0065], [0075], [0086]. Ocampo explains that "if the user attributes indicate that a user is located close to the user device and that the user uttered a command in a whispering tone," i.e., classified as a "first voice," it will adjust the user device's output accordingly based on this classification. Ocampo, [0058],

[0075], [0086]. This allows Ocampo's device to provide different outputs in response to the user command based on classification of the user's voice as whispering near the microphone (i.e., a "first voice") or shouting from farther away (i.e., a "second voice"). Ocampo, [0011], [0032], [0062]-[0065], FIGS. 1A-1B, 2A-2B, 5.

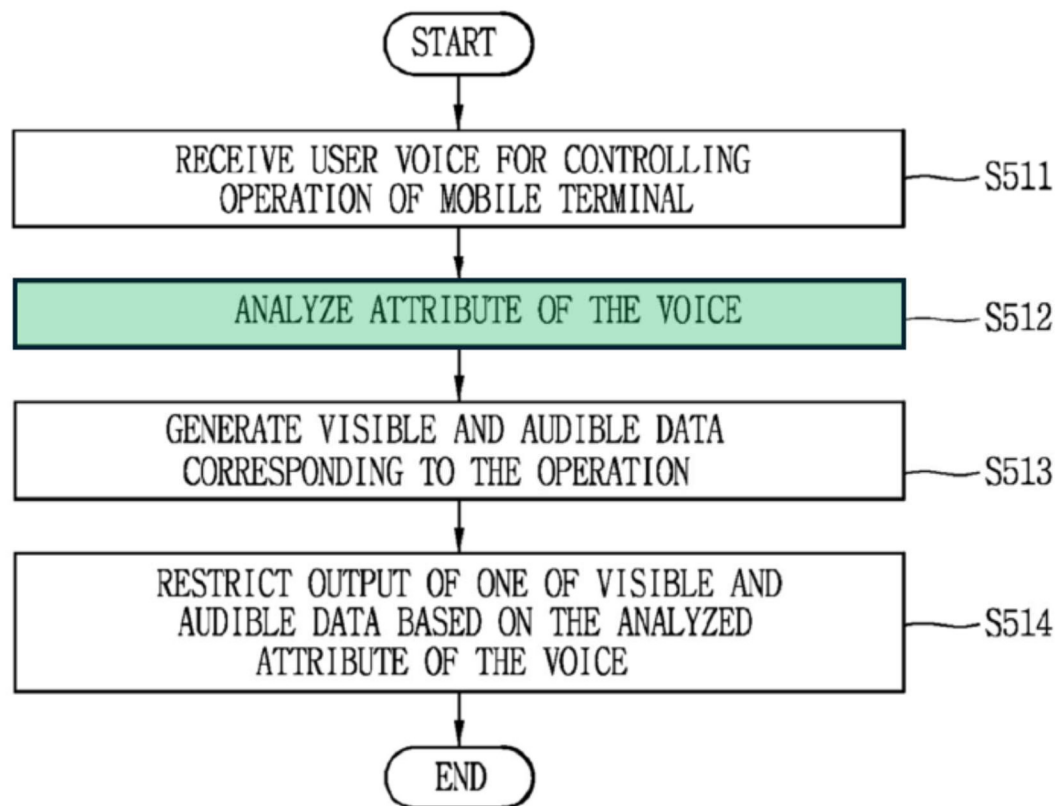
FIG. 2A



Ocampo, FIG. 2A ("first voice" classification response).

83. Yi discloses element [1b] through, for example, its explanation that “mobile terminal may further include an analyzing unit which analyzes an attribute of a voice input through the microphone 122.” Yi, [0167]. The analyzing unit analyzes attributes of the user’s voice “based on various criteria,” including distance. Yi, [0167], [0182]-[0184], FIGS. 4 (S512), 5B (S512a-S512c).

FIG. 4



Yi, FIG. 4 (annotated).

84. As Yi explains, “when the user is not sensed adjacent to the microphone 122” (first voice), the controller will “control the output unit to output the audible data and the visible data.” Yi, [0184]. But “[w]hen the microphone 122 is sensed to be located close to the user” (second voice), the controller will “control the audio output module 153 to restrict the output of the audible data.” Yi, [0184]. Thus, Yi also discloses [1b] because its mobile terminal classifies the user’s voice based on distance from the user.

d) [1c] a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user;

85. Ocampo discloses [1c] because it describes a processor to process the user’s voice to receive a user command, determine the appropriate application to handle the received command, and retrieve the necessary data to provide a tailored audio output responding to the command. *See, e.g.*, Ocampo, [0023]-[0025], [0047], [0051]-[0053]. Ocampo explains that its user device includes a processor 510 with various components and classifiers to execute required processing. Ocampo, [0047], [0066], FIG. 5.

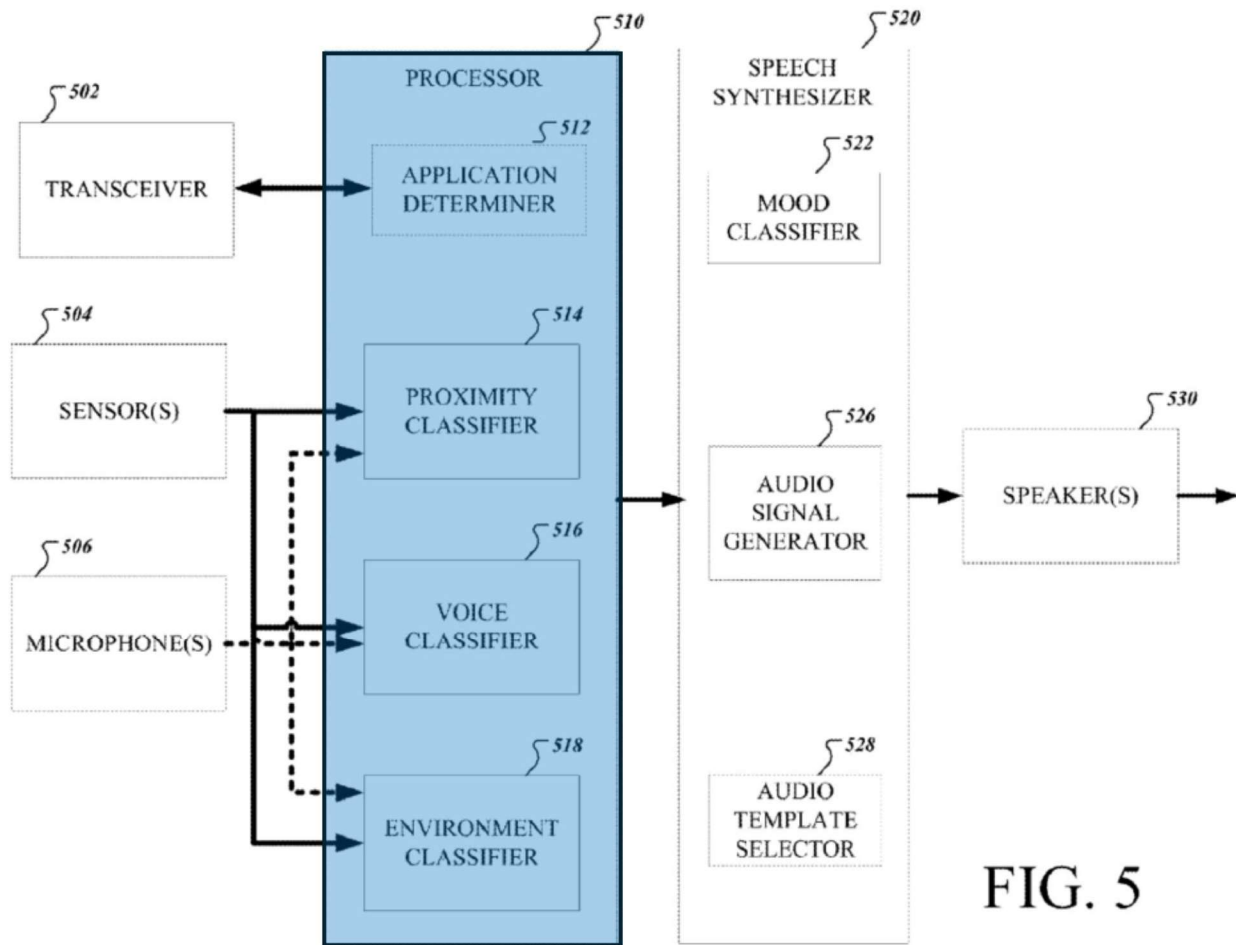


FIG. 5

Ocampo, FIG. 5 (annotated).

86. Processor 510 is responsible for executing Ocampo's disclosed methods, including analyzing the voice acquired by microphones 506 and determining the processing necessary to respond with an answer. Ocampo, [0047], FIGS. 4, 5. Upon receiving a command, the application determiner 512 identifies which application should process or respond to the command. Ocampo, [0051]. It makes that identification by categorizing the command as belonging to a classification mapped to specific applications, such as multimedia, text messaging,

or browser applications. Ocampo, [0051]. The mapping of commands to applications can be predefined by the device manufacturer, a program writer, or specified by the user. Ocampo, [0052]. Upon determining the appropriate application, Ocampo's system retrieves the necessary data for processing and otherwise responds to the command. Ocampo, [0053], [0054]. "The data 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]. Thus, responsive data can be obtained from various sources, including networks, servers, databases, or storage devices, depending on the application and command type. Ocampo, [0056].

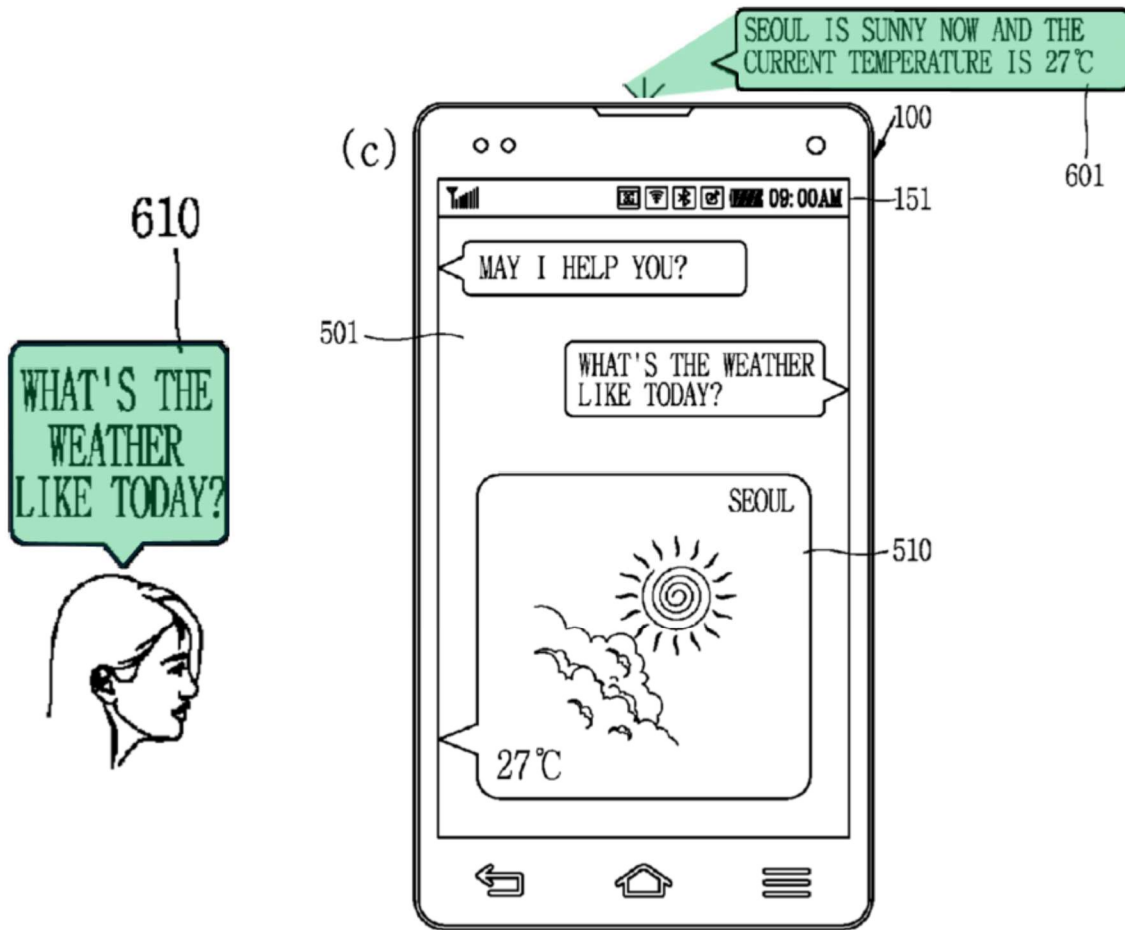
- e) **[1d] a voice-content generating unit configured to generate, based on content of the processing executed by the process executing unit, output sentence that is text data for a voice to be output to the user; and**

87. Ocampo and Yi disclose [1d]. For example, Ocampo discloses [1d] because it describes application determiner 512 selecting specific applications to generate content responsive to a user command using TTS (text-to-speech) output. Ocampo, [0047]-[0053]. "Next, the audio signal that includes the obtained data in an audio format is output using one or more speakers 530 (416)." Ocampo, [0078]. Ocampo also explains that the user device "may use any suitable audio synthesizer

technique” to convert the retrieved text data to an audio signal. Ocampo, [0077].

Ocampo thus describes generating the text data used for voice output based on processed content according to element [1d].

88. Yi also describes controller 180 generating visible and audible data responsive to a user command. Yi, [0164]. For example, Yi explains that where a user asks, “What’s the weather like today?” controller 180 “execute[s] a function for providing information related to today’s weather” and “generate[s] data to be provided to the user among the [responsive] information.” Yi, [0164], FIG. 5A(c).



Yi, FIGS. 5A(a) (left) (excerpt) (annotated), 5A(c) (right) (annotated).

89. Yi also explains that “[t]he audible data may correspond to a sound signal generated by converting at least part of the visible data into a voice.” Yi, [0166]. “For example, the controller 180 may generate the audible data by converting information corresponding to the user’s voice among the text of the visible data.” Yi, [0166]. Thus, Yi also discloses element [1d].

f) [1e] an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto, wherein

90. Ocampo and Yi disclose [1e]. Ocampo discloses this element through its dynamic text-to-speech (TTS) provisioning and use of audio template selector 528 to adjust the volume of responses spoken to a user. Ocampo, [0004], [0024], [0027], [0073], FIGS. 1A, 1B, 4 (S412, S414, S416), 5. “[T]he user device may determine how far the user is from the user device and adjust a volume or intensity of the audio output signal accordingly.” Ocampo, [0004], [0073]. As Ocampo explains, “the determined user attributes and environment attributes may be used by the audio template selector 528 to select an audio template for an audio output signal (412).” Ocampo, [0073]. In this way, when the user attributes indicate that a user is located close to the user device, “the audio template selector 528 in the user device may select an audio output template that has a low output volume and a whispering tone.” Ocampo, [0073], [0075].

91. Yi also discloses the claimed output controller with its audio output module 153 for outputting visible and audible data responsive to user commands. Yi, [0165]. “The audible data may correspond to a sound signal generated by converting at least part of the visible data into a voice.” Yi, [0166]. Said another way, “the controller 180 may generate the audible data by converting information corresponding to the user’s voice among the text of the visible data.” Yi, [0166].

Yi further explains that “the output of the audible data may be restricted by controlling the volume of the voice.” Yi, [0221].

g) [1f] the voice-content generating unit is further configured to

i. [1f.i] generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice, and

92. Ocampo discloses [1f.i] because it describes an application determiner 512 selecting an application to generate TTS (text-to-speech) output data responsive to a user command based on voice characteristics. Ocampo, [0018], [0024], [0037], [0038], [0047]-[0053], [0057], [0065], [0073], [0075], FIGS. 1A, 2A-2B, 4, 5.

93. Upon receiving a user command for information, Ocampo’s application determiner 512 retrieves responsive information and generates a response based on user attributes determined from an analysis of the user’s proximity and voice characteristics. Ocampo, Abstract, [0051], [0057]. For example, when the user device determines that the user is not near the device or is shouting the command (i.e., classifies the acquired voice as a first voice), the system will generate a text-based response for vocalizing at a louder volume. Ocampo, [0027], [0073], FIGS. 1B, 2B.

FIG. 1B

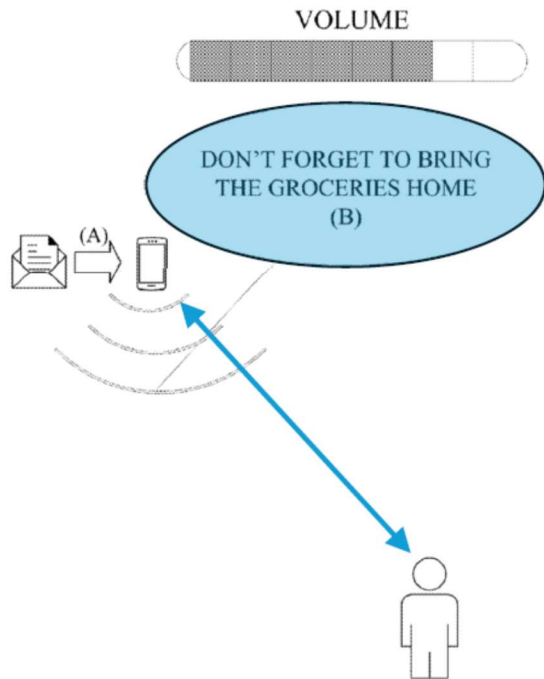


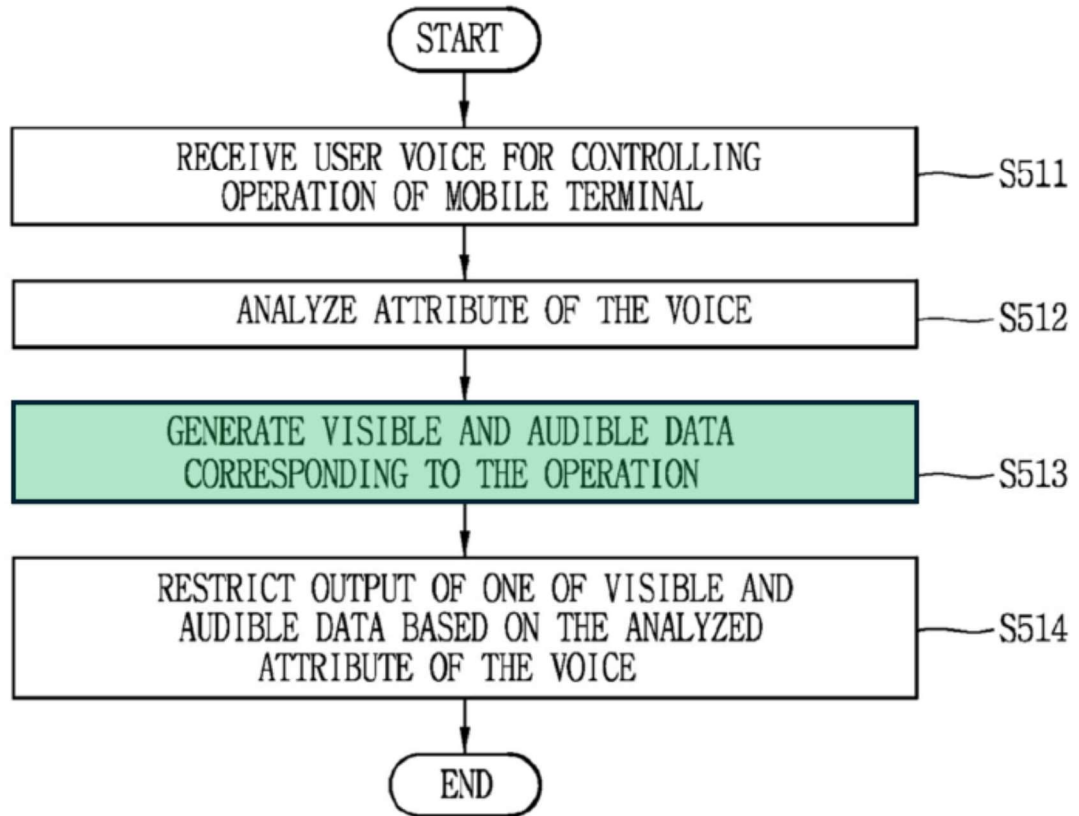
FIG. 2B



Ocampo, FIGS. 1B (left) (annotated), 2B (right) (annotated).

94. Likewise, Yi discloses [1f.i] because it discloses generating text to vocalize in response to user commands based on an analysis of the user's voice. Yi, [0167], [0170]-[0171], [0182], [0183], [0253]. Yi explains that its mobile terminal will analyze the voice received at its microphone to identify attributes of the voice based on various criteria, including the user's proximity when voicing the command, whether the user whispered, etc. Yi, [0167], [0170]-[0171], [0182], [0183], [0253]. And based on the voice classification, Yi's controller 180 generates visible and audible data responsive to the user command. Yi, [0164], FIG. 4 (S513).

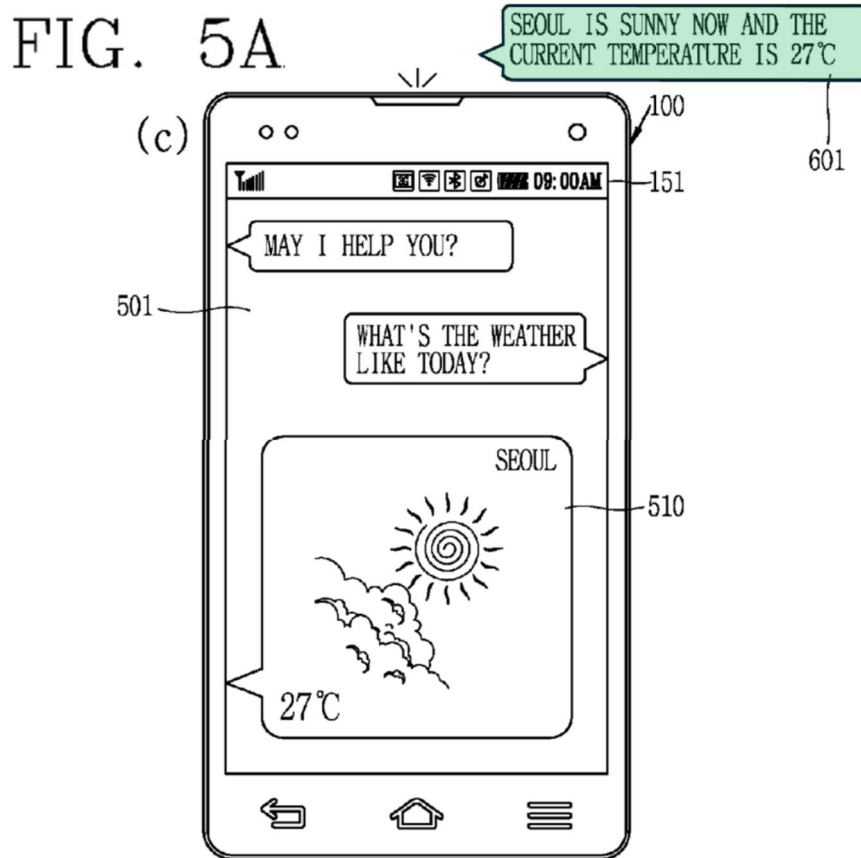
FIG. 4



Yi, FIG. 4 (annotated).

95. Yi provides specific, representative examples. Where the user asks, “What’s the weather like today?” Yi explains that controller 180 will “execute a function for providing information related to today’s weather” that accounts for the user’s proximity when “generat[ing] data to be provided to the user among the [responsive] information.” Yi, [0163]-[0167], [0188], FIG. 5A(c). In particular, when the user is distanced from the microphone while asking, “what’s the weather like today?” (i.e., classifies the user’s voice as a first voice), the mobile terminal

will respond audibly with a first output sentence: “Seoul is sunny now and the current temperature is 27°C.” Yi, FIG. 5A(c) and accompanying description.



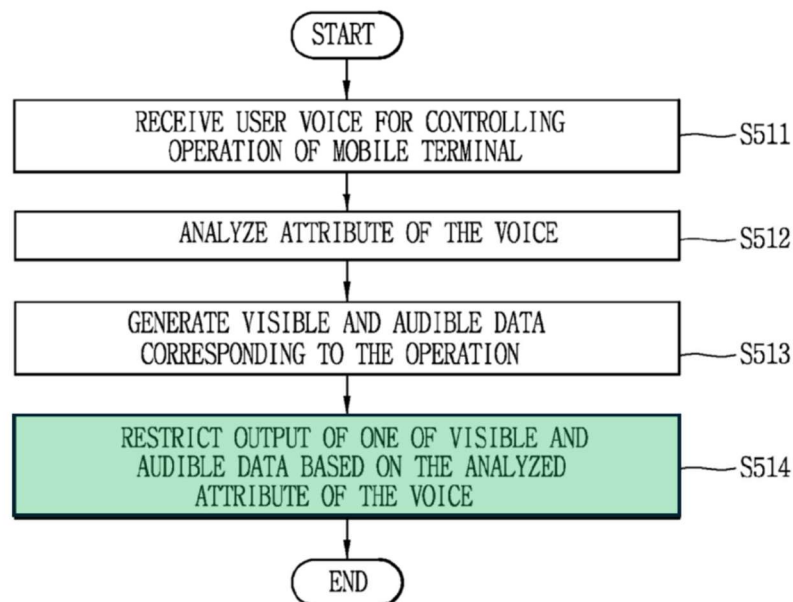
Yi, FIG. 5A(c) (annotated).

- ii. **[1f.ii] generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice, wherein**

96. In my opinion, one of skill in the art would have found element [1f.ii] obvious considering Ocampo in combination with Yi. For example, Yi discloses this element by its teaching to omit information from generated responses to a user

command based on the user's distance from the device. Yi, [0167], [0170]-[0171], [0182]-[0184], [0253], FIG. 4 (S514). As I discussed above regarding generation of the "first output sentence" (element 1f.i), Yi's mobile terminal analyzes voice input through its microphone based on various criteria for voice classification, including whisper status and proximity to the microphone. Yi, [0012], [0167], [0170]-[0171], [0182], [0183], [0253], FIG. 4 (S512, S513). Yi further explains that its mobile terminal "may control the output unit to restrict (limit) an output of at least one of the visible data and the audible data based on the voice (S514)." Yi, [0167], [0184], FIG. 4 (S514).

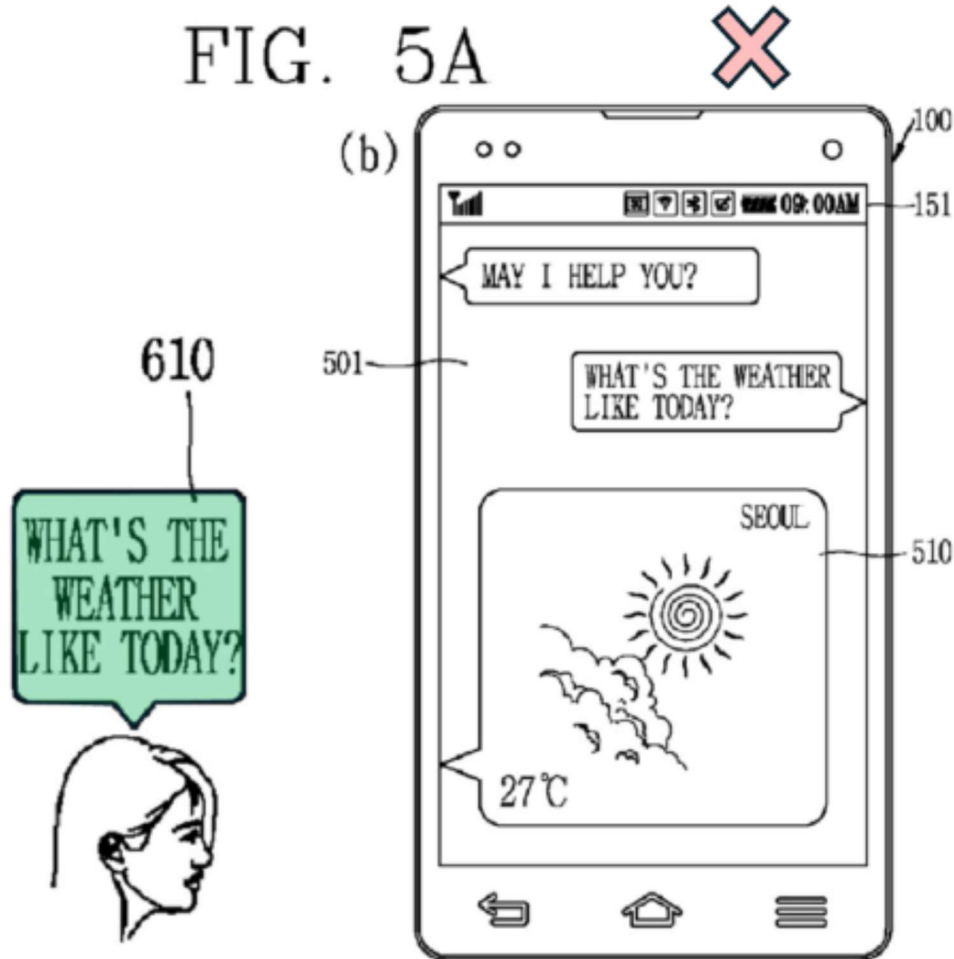
FIG. 4



Yi, FIG. 4 (annotated).

97. In particular, “[w]hen the microphone 122 is sensed to be located close to the user” (i.e., classifies the user’s voice as a second voice), “the controller 180 may control the audio output module 153 to restrict the output of the audible data (S514’).” Yi, [0184]. “That is, when the user consciously puts the mobile terminal close to the mouth to input the voice and when the mobile terminal and the user are closely located due to a surrounding environment, the output of the audible data may be restricted” to generate a “second output sentence” omitting responsive information. Yi, [0187]-[0188]. This way, “when the user inputs the voice by putting the mobile terminal close to the mouth so as not to be heard by others, the audible data may not be exposed to the outside as well.” Yi, [0188].

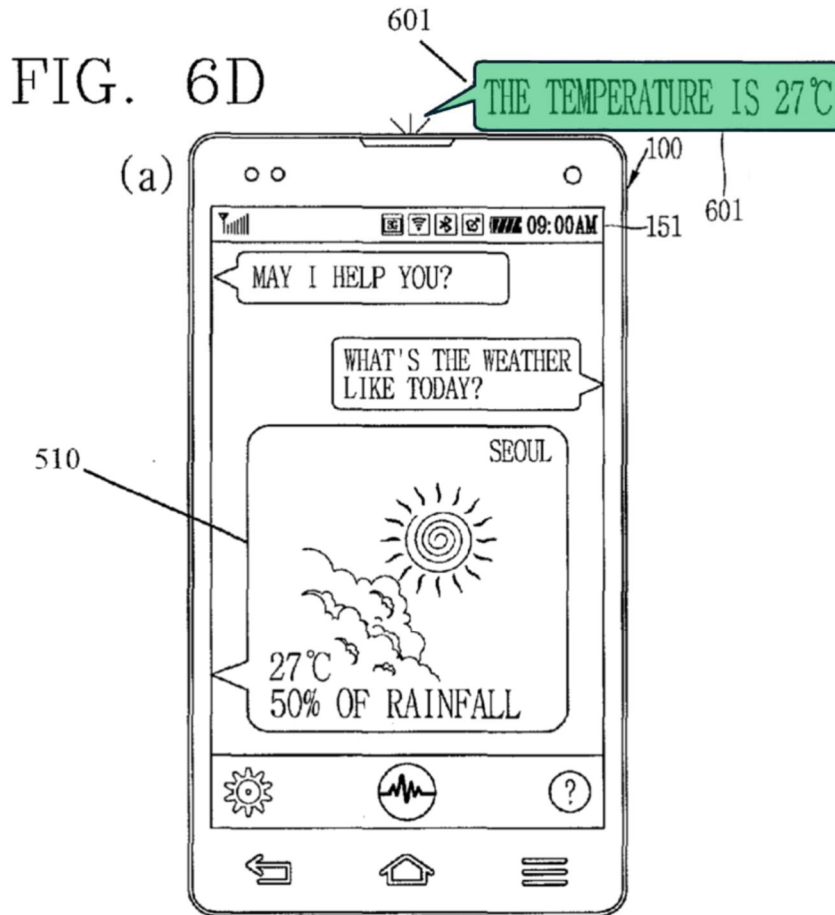
98. Yi provides figures with this omission of textual information from the generated response as compared to a “first output sentence” using the same weather inquiry as the example. Yi, [0196], FIGS. 5A, 6A-6D, 8, 9A-9B. Yi’s mobile terminal omits information when responding about the weather when the user is sensed close to the microphone. Yi, [0184], [0188], FIGS. 5A(b), 5B (S512c, S514’). For example, instead of audibly responding “Seoul is sunny now and the current temperature is 27°C,” Yi describes omitting some or all of the audible response. Yi, [0184], [0188], FIGS. 5A(b), 5B (S512c, S514’). For example, Figure 5A(c) of Yi illustrates omitting the vocalized response to the user entirely and providing only a displayed response:



Yi, FIGS. 5A(a) (left) (excerpt) (annotated), 5A(b) (second output sentence) (annotated).

99. Yi also explains that “[t]he audible data 601 may include at least part of information included in the visible data 510.” Yi, [0186], FIGS. 5-6. Yi then goes on to disclose other audible responses to the same request (i.e., “What’s the weather like today?”) that omit a subset of the audible information elsewhere vocalized in response. Yi, [0206], FIG. 6D(a). For example, Yi also illustrates the

mobile terminal omitting the geographic location and “sunny” characterization from audible data:



Yi, FIG. 6D(a) (annotated).

100. Thus, it is my opinion Yi discloses [1f.ii] because it describes generating a “second output sentence” for a given command that omits information when the user is near the device as compared to a “first output sentence” when the user is distanced from the device.

101. Moreover, it is my opinion a POSITA would have been motivated to

apply Yi's techniques for restricting information vocalized to a user based on proximity and tone of voice into Ocampo's system that already adjusted the volume of its responses based on the same user attributes. One of ordinary skill would appreciate Yi's explanation that restricting the audible data outputted "prevent[s] data from being transferred even to another user" in quiet settings where the user has whispered a command. Yi, [0170], [0180]. One of skill would also know of Yi's recognition that "a user transfers a voice in [the] form of whispering when the user talks to a person close to him/her." Yi, [0170], [0180].

102. Implementing Yi's techniques for discreetly voicing responses to a user command by omitting information would have reflected the mere combination of known elements according to known methods to yield predictable results. As explained above, both references use ubiquitous speech synthesis techniques (e.g., text-to-speech (TTS)) to achieve predictable results (e.g., voicing the text as modified for response.). *See, e.g.*, Ocampo, [0001]-[0003]; Yi, [0166], [0235]. Indeed, both references recognize the importance of maintaining user privacy; Ocampo, [0071]; Yi, [0170], [0180].

103. Thus, it is my opinion a POSITA would have had the skill and motivation to combine Ocampo's known techniques for generating TTS outputs in response to a user command with Yi's known means for restricting output to minimize exposure to third parties.

- h) [1g] the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto.**

104. Both Ocampo and Yi disclose [1g]. As I discussed above, Ocampo discloses adjusting the volume or intensity of an audio output signal based on user attributes when responding. Ocampo, [0004], [0073]. Ocampo explains that the user device converts the responsive textual data into an audio signal and controls the volume of that audio signal proportional to the determined proximity indicator. Ocampo, [0024], FIGS. 1A-1B. But when the user attributes indicate that a user is located close to the user device, “the audio template selector 528 in the user device may select an audio output template that has a low output volume and a whispering tone.” Ocampo, [0073], [0075].

105. Yi likewise discloses the claimed output controller because it describes how controller 180 uses audio output module 153 to output visible and audible data in response to user requests. Yi, [0165]. “The audible data may correspond to a sound signal generated by converting at least part of the visible data into a voice.” Yi, [0166]. Said another way, “the controller 180 may generate the audible data by converting information corresponding to the user’s voice among the text of the visible data.” Yi, [0166]. Yi further explains that “the output

of the audible data may be restricted by controlling the volume of the voice.” Yi, [0221].

5. Claim 2

- a) **[2a]: The voice-content control device according to claim 1, wherein the process executing unit comprises: an intention analyzing unit configured to extract intention information indicating an intention of the user based on the voice acquired by the voice acquiring unit; and**

106. Ocampo discloses [2a] because it describes processor 510 having a linguistic classifier that recognizes, among other things, the user’s intention from spoken commands received at its microphone. Ocampo, [0022], [0025], [0047], [0049]-[0052], [0066]. “[T]he linguistic classifier may identify words in the received audio signal,” and, based on the recognized speech, “determine which application to use to process or respond to the command and whether the determined application is configured for TTS output.” Ocampo, [0051], [0066]. For example, when Ocampo’s user device receives the voiced request, “Can you remind me what’s on my to-do list?” the device determines the user’s intention to consult “the user’s to-do or reminder list to respond to the user query.” Ocampo, [0029], [0036], FIG. 2A.

b) [2b]: an acquisition content information acquiring unit configured to acquire acquisition content information which is notified to the user based on the extracted intention information, and

107. Ocampo also discloses [2b] because it describes using application determiner 512 to acquire information responsive to spoken user commands based on the identified intention of the user. Ocampo, [0049], [0051]-[0053], [0056]. Ocampo explains that once application determiner 512 classifies the command according to the user's voiced intention, "[t]he user device may obtain data from any suitable source to respond to the user query." Ocampo, [0036], [0049]-[0053], [0056].

108. As Ocampo further explains, "[t]he data 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]. Moreover, "[t]he source from where data is obtained from depends on various factors[,] including the type of application and type of command." Ocampo, [0056]. The acquired information responsive to the user command is then converted to voice using dynamic TTS provisioning to output the response as an audio signal to the user. Ocampo, [0049], [0077]-[0078], FIG. 4 (414, 416). Ocampo therefore discloses that the acquired content gets notified to the user as an "audio signal that includes the obtained data in an audio

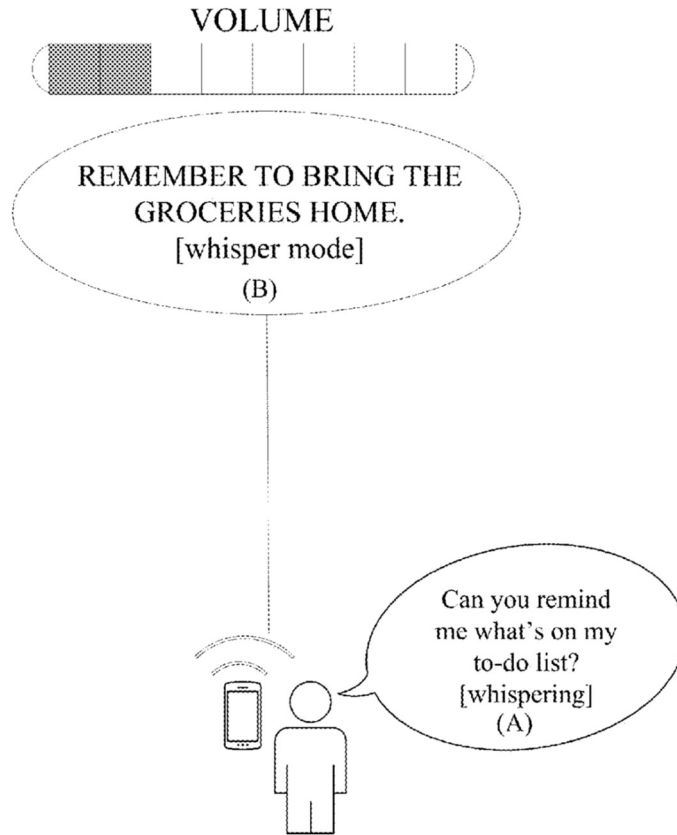
format . . . output using one or more speakers.” Ocampo, [0078], FIG. 4 (416).

- c) **[2c]: the voice-content generating unit is further configured to generate the text data including the acquisition content information as the output sentence.**

109. Ocampo discloses [2c] because it describes using dynamic text-to-speech (TTS) provisioning in order to generate the text data voiced to the user based on the acquired content responsive to a given user command. Ocampo, [0049]. I also discuss this above as it relates to element [1d]’s “voice-content generating unit configured to generate, based on content of the processing executed by the process executing unit, [an] output sentence that is text data for a voice to be output to the user.” In my opinion, one of ordinary skilled in the art would thus understand that TTS provisioning involves converting written text generated in response to a user command into spoken words using synthetic voices.

110. As Ocampo depicts in Figure 2A, when receiving the voiced request, “Can you remind me what’s on my to-do list?” the user device will consult “the user’s to-do or reminder list to respond to the user query.” Ocampo, [0036], FIG. 2A. Based on the acquired content, the user device generates a text response vocalized to the user using TTS provisioning:

FIG. 2A



Ocampo, FIG. 2A.

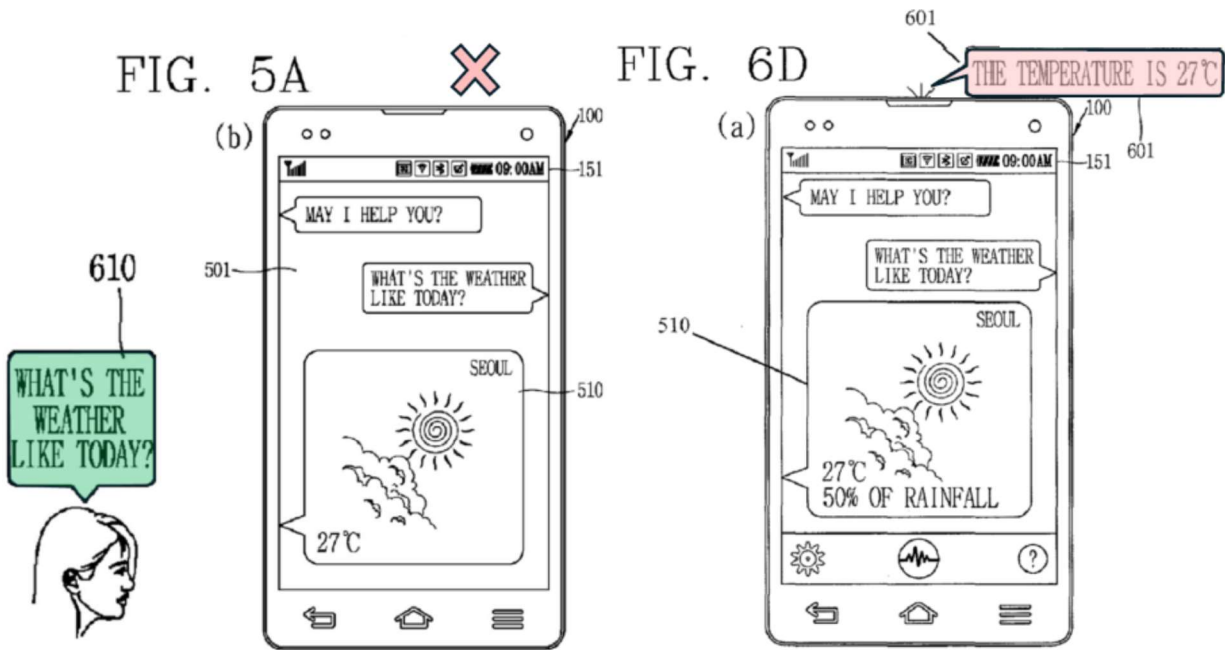
111. “As shown in FIG. 2A, the user device outputs an audio signal to inform the user that bringing the groceries home was on the user’s to-do list (B).”

Ocampo, [0037]; *see also* Ocampo, [0029], [0036].

6. **Claim 3: The voice-content control device according to claim 1, wherein, on generating the second sentence, the voice-content generating unit is further configured to omit a part of information included in the voice spoken by the user.**

112. Yi discloses claim 3 by its description of omitting information from

the voice spoken by the user when generating a “second [output] sentence” based on the acquired voice being classified as the “second voice.” As I discussed for element [1f.ii], Yi generates a second sentence omitting information by “control[ing] the output unit to restrict (limit) an output of at least one of the visible data and the audible data based on the voice (S514).” Yi, [0167], FIG. 4 (S514). Yi further discloses omitting a part of information included in the voice spoken by the user. For example, Figures 5A and 6D of Yi illustrate the mobile terminal avoiding redundant information (e.g., “weather” and “today”) from the voice spoken by the user when providing its response to the user’s weather inquiry:



Yi, FIGS. 5A(a) (left) (excerpt) (annotated), 5A(b) (center) (annotated), 6D(a) (right) (annotated)

7. Independent Claim 4

a) [4pre] A voice-content control method, comprising:

113. To the extent the preamble is limiting, Ocampo discloses it by describing various methods for controlling a user device's voiced response to a user command. Ocampo, [0003], [0024], [0028]. Ocampo discloses a voice-content control device that automatically controls and modifies audio responses to user commands based on multiple factors, including the user's proximity to the device. Ocampo, Abstract, [0003]. For example, "[i]n response to receiving a command to provide information to a user, a device retrieves information and determines user and environment attributes including: (i) a distance between the device and the user when the user uttered the query; and (ii) voice features of the user." Ocampo, Abstract. Indeed, Ocampo explains that the "TTS operation executed on a user device may automatically control and modify an audio output based on multiple factors[,] including the user's voice" Ocampo, [0003].

b) [4a] calculating a distance between a user and a voice-content control device;

114. Ocampo discloses [4a] for reasons I discussed regarding [1a] above.

c) [4b] acquiring a voice spoken by a user;

115. Ocampo discloses [4b], as I similarly discussed above regarding element [1b] regarding "acquired by a voice acquiring unit"). Ocampo explains

that the user device actuates its microphones to obtain “samples of the user’s voice.” Ocampo, [0031]. The user device uses these samples to determine user attributes. Ocampo, [0026].

- d) **[4c] analyzing the acquired voice to classify the acquired voice as either one of a first voice and a second voice based on the distance between the user and the voice-content control device;**

116. Ocampo discloses [4c] for reasons I discussed regarding [1b] above.

- e) **[4d] analyzing the acquired voice to execute processing intended by the user;**

117. Ocampo discloses [4d] for reasons I discussed regarding [1c] above.

- f) **[4e] generating, based on content of the executed processing, [an] output sentence that is text data for a voice to be output to the user; and**

118. Ocampo discloses [4e] for reasons I discussed regarding [1d] above.

- g) **[4f] adjusting a sound volume of voice data obtained by converting the output sentence thereinto, wherein at the generating,**

119. Ocampo discloses [4f] for reasons I discussed regarding [1e] above.

- i. **[4f.i] a first output sentence is generated as the output sentence when the acquired voice has been classified as the first voice, and**

120. Ocampo discloses [4f.i] for reasons I discussed regarding [1f.i] above.

- ii. **[4f.ii] a second output sentence is generated as the output sentence in which a part of information included in the first output sentence is omitted when the acquired voice has been classified as the second voice, wherein**

121. Ocampo discloses [4f.ii] for reasons I discussed regarding [1f.ii] above.

- h) **[4g] at adjusting the sound volume of voice data, further adjusting the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of voice data obtained by converting the second output sentence thereinto.**

122. Ocampo discloses [4g] for reasons I discussed regarding [1g] above.

8. Independent Claim 5

- a) **[5pre] A non-transitory storage medium that stores a voice-content control program that causes a computer to execute:**

123. To the extent the preamble is limiting, Ocampo discloses it by explaining that “all of the functional operations and/or actions described in this specification” may be implemented as computer-program instructions “encoded on a computer readable medium for execution by, or to control the operation of, [a] data processing apparatus.” Ocampo, [0090]. “The computer-readable medium may be a machine-readable storage device, a machine-readable storage substrate, a memory device, a composition of matter effecting a machine-readable propagated

signal, or a combination of one or more of them.” Ocampo, [0090].

b) [5a] calculating a distance between a user and a voice-content control device;

124. Ocampo discloses [5a] for reasons I discussed regarding [1a] above.

c) [5b] acquiring a voice spoken by a user;

125. Ocampo also discloses [5b] for the reasons discussed above regarding element [1b]’s “acquired by a voice acquiring unit.” Ocampo explains that the user device actuates its microphones to obtain “samples of the user’s voice.” Ocampo, [0031]. The user device uses these samples to determine user attributes. Ocampo, [0026].

d) [5c] analyzing the acquired voice to classify the acquired voice as either one of a first voice and a second voice based on the distance between the user and the voice-content control device;

126. Ocampo discloses [5c] for reasons I discussed regarding [1b] above.

e) [5d] analyzing the acquired voice to execute processing intended by the user;

127. Ocampo discloses [5d] for reasons I discussed regarding [1c] above.

f) [5e] generating, based on content of the executed processing, [an] output sentence that is text data for a voice to be output to the user; and

128. Ocampo discloses [5e] for reasons I discussed regarding [1d] above.

- g) [5f] adjusting a sound volume of voice data obtained by converting the output sentence therein, wherein at the generating,**

129. Ocampo discloses [5f] for reasons I discussed regarding [1e] above.

- i. [5f.i] a first output sentence is generated as the output sentence when the acquired voice has been classified as the first voice, and**

130. Ocampo discloses [5f.i] for reasons I discussed regarding [1f.i] above.

- ii. [5f.ii] a second output sentence is generated as the output sentence in which a part of information included in the first output sentence is omitted when the acquired voice has been classified as the second voice wherein**

131. Ocampo discloses [5f.ii] for reasons I discussed regarding [1f.ii]

above.

- h) [5g] at adjusting the sound volume of voice data, further adjusting the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence therein differs from the sound volume of voice data obtained by converting the second output sentence therein.**

132. Ocampo discloses [5g] for reasons I discussed regarding [1g] above.

XII. CONCLUSION

133. This declaration and my opinions herein are made to the best of my knowledge and understanding, and based on the material available to me, at the time of signing this declaration. I declare that all statements made herein on my

Declaration of Mr. Stuart Lipoff
Inter Partes Review of U.S. Patent No. 11,069,337

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

Date: 6/12/25



Stuart Lipoff