

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

**AMAZON.COM, INC.,
AMAZON.COM SERVICES LLC, and
AMAZON WEB SERVICES, INC.,**
Petitioners,

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner.

Case No. IPR2024-00673
Patent No. 11,244,675

DECLARATION OF RICHARD STERN, PH.D.

<i>Amazon v. SoundClear</i> US Patent 11,244,675 Amazon Ex. 1002
--

TABLE OF CONTENTS

I.	BACKGROUND -----	1
A.	Experience and Qualifications -----	1
B.	Materials Considered-----	3
II.	APPLICABLE LEGAL STANDARDS -----	4
A.	Claim Construction -----	5
B.	Obviousness -----	6
III.	PERSON OF ORDINARY SKILL IN THE ART -----	10
IV.	TECHNOLOGY BACKGROUND -----	11
A.	Classifying Users Based on Proximity -----	12
B.	Tailoring Content Based on Distance -----	14
V.	THE '675 PATENT -----	20
A.	Overview -----	20
B.	Priority Date -----	22
VI.	CLAIMS 1-7 OF THE '675 PATENT WOULD HAVE BEEN OBVIOUS -----	22
A.	Claims 1-7 Would Have Been Obvious in View of Shin and Aoyama -----	22
1.	Claim 1 -----	23
a.	1[pre]: Output-Content Control Device -----	23
b.	1[a]: Voice Acquiring Unit -----	24
c.	1[b]: Voice Classifying Unit -----	26

i.	1[b][i]: Calculates Distance -----	26
ii.	1[b][ii]: Classifies Voice Based on Distance-----	28
d.	1[c]-1[d]: Intention Analyzing Unit and Notification-Information Acquiring Unit -----	32
e.	1[e]: Output-Content Generating Unit-----	39
f.	1[f]: Generate a First Output Sentence -----	41
g.	1[g]: Generate a Second Output Sentence -----	47
2.	Claim 2 -----	49
a.	2[a]: Notification Information -----	50
b.	2[b]: Output-Content Generating Unit-----	52
3.	Claim 3 -----	53
4.	Claim 4 -----	56
a.	4[a]: Relationship Information-----	56
b.	4[b]: Output-Content Generating Unit-----	57
5.	Claim 5 -----	58
6.	Claims 6 and 7-----	61
B.	Claims 1-2 and 5-7 Would Have Been Obvious in View of Shin and Iwase-----	64
1.	Claim 1 -----	64
2.	Claim 2 -----	68
a.	2[a]: Notification Information -----	68
b.	2[b]: Output-Content Generating Unit-----	69

3.	Claims 5-7-----	70
C.	Claims 1-7 Would Have Been Obvious in View of Shimomura and Aoyama -----	70
1.	Claim 1 -----	70
a.	1[pre]: Output-Content Control Device -----	70
b.	1[a]: Voice Acquiring Unit-----	72
c.	1[b]: Voice Classifying Unit -----	73
i.	1[b][i]: Calculates Distance -----	73
ii.	1[b][ii]: Classifies Voice Based on Distance-----	74
d.	1[c]-1[d]: Intention Analyzing Unit and Notification-Information Acquiring Unit -----	75
e.	1[e]: Output-Content Generating Unit-----	78
f.	1[f]: Generate a First Output Sentence -----	80
g.	1[g]: Generate a Second Output Sentence -----	83
2.	Claims 2-4-----	84
3.	Claim 5 -----	85
4.	Claims 6 and 7-----	87
D.	Claims 1-2 and 5-7 Would Have Been Obvious in View of Shimomura and Iwase -----	87
1.	Claim 1 -----	87
2.	Claim 2 -----	89
3.	Claims 5-7-----	90

E.	Claims 3-4 Would Have Been Obvious in View of (a) Shin, Iwase, and Aoyama; or (b) Shimomura, Iwase, and Aoyama-----	90
1.	Claim 3 -----	90
2.	Claim 4 -----	92
F.	Claim 5 Would Have Been Obvious in View of the Previously Discussed Combinations and Schuster -----	93
G.	Claims 1-7 Would Have Been Obvious in View of the Previously Discussed Combinations and Kristjansson -----	96
VII.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	99
VIII.	CONCLUSION -----	99

I, Richard Stern, Ph.D., do hereby declare:

1. I am making this declaration at the request of Petitioners Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively, “Amazon”). I have been retained by Amazon as a technical expert in this matter.

2. I am being compensated for my work on this case. My compensation does not depend on the content of this Declaration or the outcome of these proceedings. As part of a diversified portfolio, I own shares of Amazon stock. A financial professional manages my portfolio, and I do not. This stock ownership constitutes less than 1 percent of the total portfolio, and does not in any way affect the contents of my opinions in this Declaration.

I. BACKGROUND

A. Experience and Qualifications

3. I am a Professor at Carnegie Mellon University in the Department of Electrical and Computer Engineering, the Department of Computer Science, and the Language Technologies Institute. I have been on the faculty of Carnegie Mellon since 1977.

4. I received the S.B. degree from the Massachusetts Institute of Technology (MIT) in 1970, the M.S. from the University of California, Berkeley, in 1972, and the Ph.D. from MIT in 1977, all in electrical engineering.

5. While I was a graduate student, I worked as a Teaching and Research Assistant in the Department of Electrical Engineering at MIT, from 1973 to 1976. My teaching experience was in the areas of signal processing and acoustics. My research at MIT had been in the area of auditory processing and perception.

6. I am a fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Acoustical Society of America, and the International Speech Communication Association (ISCA). I was the ISCA 2008-2009 Distinguished Lecturer, a recipient of the Allen Newell Award for Research Excellence in 1992, and I served as the General Chair of Interspeech 2006. Interspeech is the world's largest technical conference focused on speech processing and application.

7. Much of my current research is in spoken language systems, where I am particularly concerned with the development of techniques with which automatic speech recognition can be made more robust with respect to changes in environment and acoustical ambience. Among my longstanding work on speech processing are several projects involving speaker identification. Furthermore, as a Professor, I have taught 15 courses, many of which are at the graduate level, in the general areas of signal processing, communication theory, and acoustics.

8. My relevant publications are available on Carnegie Mellon's web site at <http://www.cs.cmu.edu/afs/cs/user/robust/www/papers.html>.

9. I understand a copy of my current curriculum vitae, which lists my publications for the last ten years, is being submitted as Exhibit 1006.

B. Materials Considered

10. In preparing this Declaration, I have considered the following materials:

Exhibit No.	Description
1001	U.S. Patent No. 11,244,675
1003	U.S. Patent App. Publ. No. 2017/0083281 (“Shin”)
1004	English Translation of Shimomura from Ex. 1005
1005	Declaration of Gwen Snorteland for Translation of Japanese Unexamined Patent App. Publ. 2005/202076 (“Shimomura”)
1007	Excerpts from File History of U.S. Patent No. 11,244,675
1009	U.S. Patent App. Publ. No. 2017/0154626 (“Kim”)
1010	U.S. Patent App. Publ. No. 2017/0337921 (“Aoyama”)
1011	U.S. Patent App. Publ. No. 2016/0284351 (“Ha”)
1012	U.S. Patent No. 9,489,172 (“Iyer”)
1013	U.S. Patent No. 10,147,439 (“Kristjansson”)
1015	Nicolae Duta, <i>Natural Language Understanding and Prediction: From Formal Grammars to Large Scale Machine Learning</i> , 131 Fundamenta Informaticae 425 (2014) (“Duta”)
1016	U.S. Patent No. 9,680,983 (“Schuster”)
1017	U.S. Patent No. 11,183,167 (“Iwase”)
1018	U.S. Patent App. Publ. No. 2019/0103127 (“Tseretopoulos”)

11. I understand that Shimomura published as a Japanese Patent Application, and that it has been translated by a certified translator. (*See* Ex. 1005.) In performing my analysis, I relied on the certified English translation provided by the translator, which I understand will be submitted as Exhibit 1004.

12. I have also relied on my education, training, and experience, and my knowledge of pertinent literature in the field of the '675 patent.

II. APPLICABLE LEGAL STANDARDS

13. I have been asked to provide my opinion as to whether the claims of the '675 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention, in view of the prior art.

14. I am an electrical engineer by training and profession. The opinions I am expressing in this report involve the application of my training and technical knowledge and experience to the evaluation of certain prior art with respect to the '675 patent.

15. Although I have been involved as a technical expert in patent matters before, I am not an expert in patent law. Therefore, the attorneys from Knobbe, Martens, Olson & Bear, LLP have provided me with guidance as to the applicable patent law in this matter. The paragraphs below express my understanding of how I must apply current principles related to patent validity to my analysis.

A. Claim Construction

16. It is my understanding that in determining whether a patent claim is obvious in view of the prior art, the Patent Office construes the claim by giving the claim terms their plain and ordinary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention in view of the intrinsic record (patent specification and file history). For the purposes of this review, and to the extent necessary, I have interpreted each claim term in accordance with its plain and ordinary meaning as it would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record. I have been instructed that the time of the invention is March 12, 2018, which I understand to be the earliest claimed priority date of the '675 patent.

17. I understand that a patent and its prosecution history are considered “intrinsic evidence” and are the most important sources for interpreting claim language in a patent. I also understand that in reading the claim, I must not import limitations from the specification into the claim terms; in other words, I must not narrow the scope of the claim terms by implicitly adding disclosed limitations that have no express basis in the claims. The prosecution history of related patents and applications can also be relevant.

18. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help

interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

19. Unless expressly stated herein, I have applied the plain and ordinary meaning of the claim terms, which I understand is the meaning that a person of ordinary skill in the art would have given to terms in March 2018 based on a review of the intrinsic evidence.

B. Obviousness

20. It is my understanding that a claim is “obvious” if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether a claim would have been obvious: (i) the scope and content of the prior art; (ii) the differences, if any, between each claim of the ’675 patent and the prior art; (iii) the level of ordinary skill in the art in March 2018; and (iv) additional considerations, if any, that indicate that the invention was obvious or not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” of non-obviousness or obviousness.

21. I also understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the pertinent art would have

known in March 2018. I understand that the hypothetical person of ordinary skill is presumed to have knowledge of all pertinent prior art references.

22. It is my understanding that something is “inherent in,” and therefore taught by, the prior art, if it necessarily flows from the explicit disclosure of the prior art. I understand that the fact that a certain result or characteristic *may be* present in the prior art is not sufficient to establish inherency. However, if the result or characteristic is *necessarily* present based upon the explicit disclosure in the prior art, it is inherent in the prior art and is therefore disclosed.

23. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if it is pertinent to the problem that the inventors were trying to solve. A reference is reasonably pertinent if it logically would have commended itself to an inventor’s attention in considering the problem at hand. If a reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an obviousness analysis. Here, all of the references relied on in my obviousness analysis below are in the same field of endeavor as the ’675 patent, e.g., controlling a device’s voice output. The references are also pertinent to a particular problem the inventor was focused on, e.g., adjusting voice output to improve user interaction.

24. It is my understanding that the law recognizes several rationales for combining references or modifying a reference to show obviousness of claimed subject matter. Some of these rationales include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- a predictable use of prior art elements according to their established functions;
- using known techniques to improve similar devices (methods, or products) in the same way;
- applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success (in which case a claim would have been obvious to try);
- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and

- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

25. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism by experts; (4) commercial success of a product covered by the patent; (5) unexpected results achieved by the claimed invention; (6) industry praise of the claimed invention; (7) deliberate copying of the invention; and (8) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was within the knowledge of a person of ordinary skill as of March 2018. I am not aware of any evidence of secondary considerations that would suggest that the claims of the ’675 patent would have been nonobvious in March 2018.

26. I understand that when assessing obviousness, using hindsight is impermissible; that is, what is known today or what was learned from the teachings of the patent should not be considered. The patent should not be used as a road map for selecting and combining items of prior art. Rather, obviousness must be considered

from the perspective of a person of ordinary skill at the time the alleged invention was made – March 2018 in this case.

27. I also understand that an obviousness analysis must consider the invention as a whole, as opposed to just a part or element of the invention. I understand this “as a whole” assessment to require showing that one of ordinary skill in the art at the time of invention, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

III. PERSON OF ORDINARY SKILL IN THE ART

28. It is my understanding that when interpreting the claims of the '675 patent and evaluating whether a claim is anticipated or would have been obvious, I must do so based on the perspective of a person of ordinary skill in the art at the relevant priority date. I have been instructed to assume for the purposes of my opinions that the relevant priority date of the '675 patent is March 12, 2018.

29. I understand that in determining the level of ordinary skill in the art, several factors are considered. Those factors may include: (i) the type of problems encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. A person of ordinary skill in the art must have the capability of understanding the scientific and engineering principles

applicable to the pertinent art.

30. Based on my review of the specification and claims of the '675 patent, it is my opinion that a person of ordinary skill in the art would have had a minimum of a bachelor's degree in computer engineering, computer science, electrical engineering, or a similar field, and approximately two years of industry or academic experience in a field related to controlling the audio output of electronic devices. Work experience could substitute for formal education and additional formal education could substitute for work experience.

31. My conclusions below that the claims of the '675 patent would have been obvious would remain the same even if the priority date, field of endeavor, or level of ordinary skill were slightly different.

32. I meet the above definition of a person of ordinary skill in the art, and did so as of March 2018. Also, I have worked with persons of ordinary skill in the art through my professional and academic experiences, and I have an understanding of their skill level around March 2018.

IV. TECHNOLOGY BACKGROUND

33. The '675 patent describes an electronic device, such as a speaker or phone, that receives and processes a user's voice input and outputs a voice response. (Ex. 1001 ('675 patent), Fig. 1, 2:64-3:4.) The patent claims priority to March 2018,

years after Apple launched Siri in its iPhones (2011), Amazon launched Alexa-enabled devices (2014), and Google launched its voice assistant (2016). The patent admits that devices that detect voice, perform processing according to the user's intent, and provide a voice output were known. (*Id.*, 1:20-27.) Public reporting confirms that both Siri- and Alexa-enabled devices performed these functions. (*E.g.*, *Amazon Just Surprised Everyone with a Crazy Speaker that Talks to You*, <https://www.theverge.com/2014/11/6/7167793/amazon-echo-speaker-announced> (Echo “will provide information, music, news, weather, and more whenever you ask for it,” “can hear users from across the room,” and can be used for “telling you the local weather” or asking “general questions and get[ting] answers from Wikipedia”); *iPhone 4S: a Siri-ously Slick, Speedy Smartphone*, <https://arstechnica.com/gadgets/2011/10/iphone-4s-a-siri-ously-slick-speedy-smartphone/> (“Siri offers iPhone 4S users a way to interact with their devices, apps, and data with natural language voice commands. Users can send and receive texts or e-mails simply by talking to Siri, find out whether there will be rain today, create new Reminders and calendar items, ask for directions, move appointments, and more.”).)

A. Classifying Users Based on Proximity

34. Many devices that classified users based on their proximity to the device were known in the art. For example, Kim, which was published in 2017, dis-

closed a device with this functionality. Kim disclosed that its device “collect[s] context information associated with the user” after the device receives a user’s voice input. (Ex. 1009 (Kim) ¶¶[0045], [0163] (“If the voice is collected, ... the processor 120 may collect context information. If a user voice signal is received, the processor 120 ... may collect user related information, using a sensor[.]”).) The device can obtain user information using a variety of sensors, but notably, it can use a “proximity sensor” to detect “distance information” (i.e. the distance between a user and the device). (*Id.* ¶[0068] (“The sensor 180 may include ... a proximity sensor According to an example embodiment, the sensor 180 may collect ... distance information from [the user]”).) The device can determine whether a user is a “first specified user” based on the “context information.” (*Id.* ¶[0169] (“the processor 120 ... may determine whether the context information meets a first specified condition The first specified condition may include a condition in which a user who asks is a first specified user.”); *see also id.* ¶[0244] (“the electronic device 1500 may use a sensor ... for identifying at least one user”).)

35. Aoyama, which also published in 2017, provides another example of an “information processing device” which classified users based on proximity. (Ex. 1010 (Aoyama), Abstract.) Aoyama discloses a device that determines a “user state” based on analysis of “image” and “sound input.” (*Id.* ¶[0114] (“the user state estimation unit 15 mainly estimates the user state on the basis of the analysis results of

an image and the sound input has been described above”).) The device can determine various “user states,” such as whether a user is “stopped,” “walking,” or “running.” (*Id.* ¶[0106].) Based on this determination, the device may classify users as “calm” or “hurried.” (*Id.*) A person of ordinary skill in the art would have understood that, in order to make the determination that a user is either “stopped” or “running,” the device necessarily must calculate its proximity to the user at sequential time intervals. Further, Aoyama notes that “the type of information used to estimate the user state is not particularly limited as long as the information is information such as an image, a sound input, or detection results by various sensors.” (*Id.* ¶[0114]; *see also id.* ¶[0107] (“the estimation method is not particularly limited as long as the user state estimation unit 15 can estimate ... a motion of the user”).)

B. Tailoring Content Based on Distance

36. Using a proximity sensor to determine a user’s distance from a device and then tailoring the information provided based on that distance has been known for decades. In 2005, Shimomura described an electronic device, such as a robot, that converses with a user. (Ex. 1004 (Shimomura), Abstract.) The device included cameras for determining the distance between the device and the user. (*Id.* ¶¶[0039], [0093].) The device classified the user’s speech based on distance (e.g., whether the user was more or less than 350 cm from the device) and then adjusted the output

content (e.g., including or omitting words) accordingly. (*Id.* ¶¶[0049]-[0050], [0057]-[0058], [0085], [0088], [0092].)

37. In 2017, Shin disclosed a device (e.g., smartphone, portable multimedia player, or home appliance) that acquires a user’s speech input via a microphone, analyzes it, and provides content as voice output via a speaker. (Ex. 1003 (Shin) ¶¶[0031], [0037].) Shin used a “distance detection module” or “proximity sensor” to “compute a distance between a user and the electronic device[.]” (*Id.* ¶¶[0066], [0153], [0053], [0075], [0077], Figs. 5A, 9.)

38. Shin’s device tailored voice input based on distance. For users within 1 meter, it outputted non-abbreviated content (e.g., “detailed content” in which the content is “described most precisely”). (*Id.* ¶¶[0051], [0088], Table 3.) For users more than 1 meter away, it outputted abbreviated content. (*Id.*) An example of this procedure in Shin’s system is shown in Figure 5A:

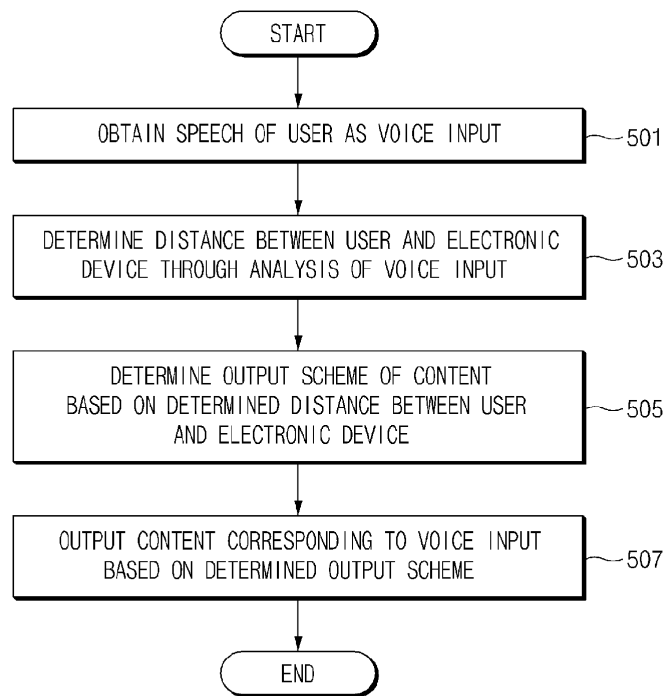


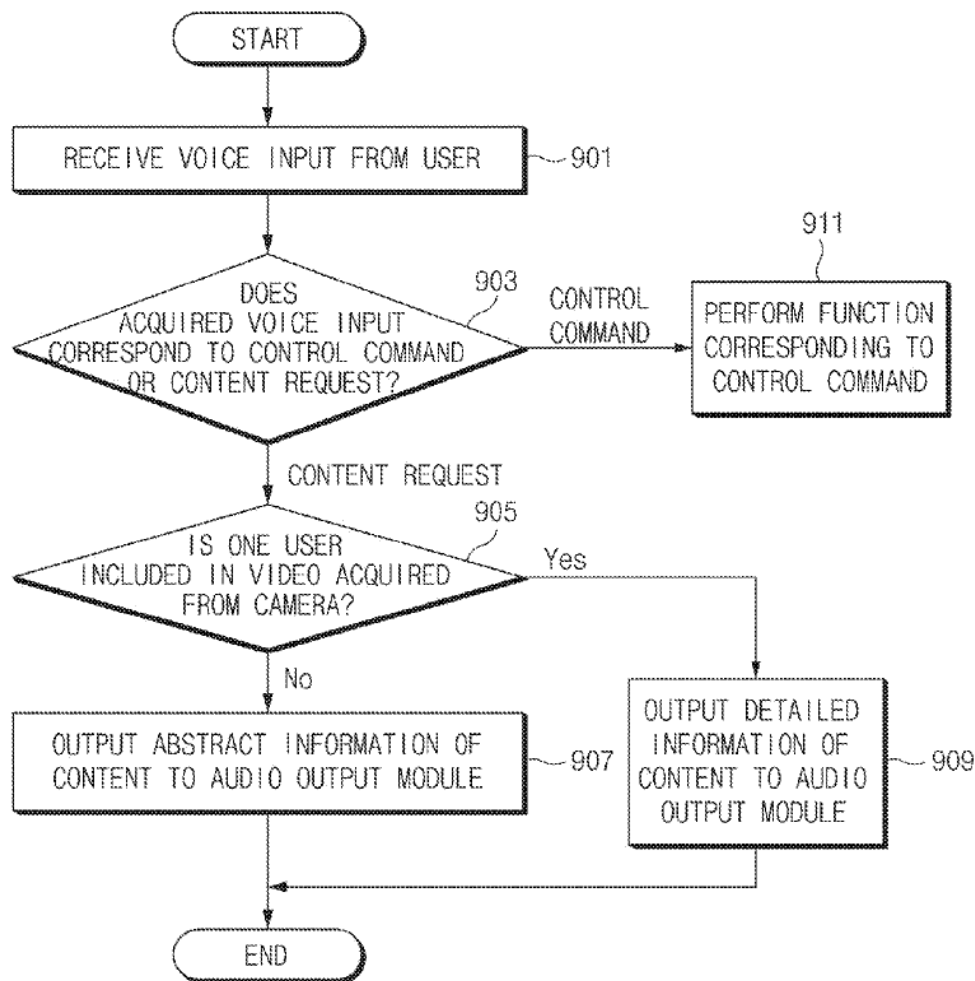
FIG. 5A

(*Id.*, Fig. 5A.)

39. As another example, if a user said, “Let me know today’s weather,” Shin’s device obtained weather-related information and responded with, for example: “The weather in Jul. 1, 2015 is (①) rainy after cloudy (②). The highest temperature is 28° C., and the lowest temperature is 18° C. (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).” (*Id.* ¶¶[0087]-[0088].) If the voice was classified as a near voice (e.g., less than 1 meter away), all five categories of information were provided. (*Id.*) If the voice was classified as farther away (e.g., user between 1 and 2 meters away), category ⑤ was omitted. (*Id.*) And, if the voice

was classified as even farther away (e.g., greater than 4 meters), all but category ② was omitted. (*Id.*) Thus, Shin’s device used a proximity sensor to calculate distance between the user and the device, classify the user’s voice based on that distance, and tailor the response’s content based on the classification. This allowed Shin’s device to provide a “suitable” amount of information at a “suitable” volume based on distance. (*Id.* ¶¶[0089], [0080].)

40. Similarly, in 2016, Ha described an electronic device that “provides content in response to [] voice input.” (Ex. 1011 (Ha), Abstract.) The device can provide either “abstract” or “detailed” information to a user. (*Id.* ¶[0042] (“the processor 120 may determine to control the audio output module 150 to output abstract information of content”); *see also id.* ¶¶[0044], [0048], [0050], [0052].) Ha’s device may determine which type of content to provide to a user based on “a video of the vicinity of the electronic device 100 shot by the camera 170.” (*Id.* ¶[0050].) Ha discloses that the device may determine if “a user is present in the vicinity of the electronic device.” (*Id.* ¶[0114].) When the device makes the determination that a user is in its vicinity, it can provide “detailed information.” (*Id.*) Ha’s Figure 9 illustrates this process:



(*Id.*, Fig. 9.) If the device’s camera registers a user after receiving voice input, it would “output detailed information.” (*Id.*) Alternatively, if the camera did not register a user (e.g. because the user was too far from the camera), or alternatively, if it registered multiple users, it would output “abstract information.” (*Id.*; *see also id.* ¶¶[0109]-[0116].) Ha’s claims provide further support for this functionality. Specifically, claim 14 states that the device comprises “a camera configured to shoot a video of a vicinity of the electronic device” and a “processor [] further configured to

determine the output scheme of the content to be outputted through the audio output module or the display based on the video of the vicinity of the electronic device.” (*Id.*, claim 14.)

41. Tailoring output by replacing words was also known. For example, Aoyama, which I discussed above (*see* ¶35, above), disclosed a mobile device that analyzes speech input, executes processing, and audibly outputs the result. (Ex. 1010 (Aoyama) ¶¶[0161], [0127], [0253].) The device customizes its output by replacing words based on the user or situation. (*Id.* ¶¶[0138]-[0140], [0177], [0218].) For example, the device may refer to a table and replace names such as “Taro Yamada” with “Mr. Yamada” in its output. (*Id.* ¶¶[0138], [0140].)

42. Iwase disclosed similar features. It described a mobile device that audibly presents notification information to users. (Ex. 1017 (Iwase), 3:26-29, 5:10-13, Abstract.) To enhance comprehension, the device adjusts the notification’s subject based on user attributes, thus tailoring its voice output by replacing certain words. (*Id.*; *id.*, 1:54-57, 4:17-29, 12:31-34, 15:6-14.) For example, the phrase “I recommend bringing an umbrella” may be modified by replacing “I recommend” with “You should,” resulting in “You should take an umbrella.” (*Id.*, 15:6-14.)

43. Similarly, Tseretopoulos (filed in 2017) disclosed an electronic device (e.g., computer or smartphone) that interprets a conversational input from a user and determines a response. (Ex. 1018 (Tseretopoulos), Abstract, ¶¶[0030], [0039].) To

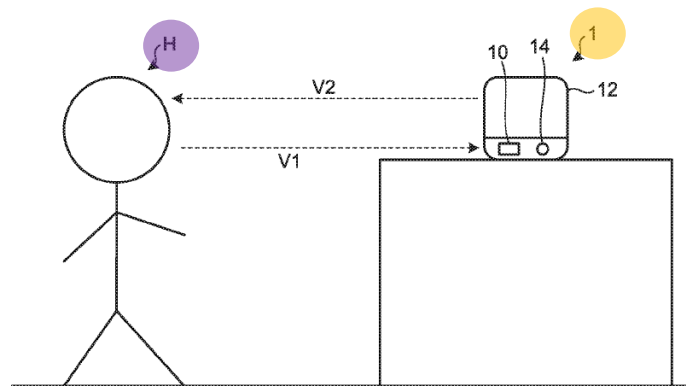
more effectively personalize the response provided to the user, the system can replace words in the response with synonyms to generate a modified version of the response that is then provided to the user. (*E.g.*, *id.* ¶[0004], Abstract.)

V. THE '675 PATENT

A. Overview

44. The '675 patent describes an “output content control device” that analyzes a user’s voice and generates a response. (Ex. 1001 ('675 patent), Abstract.) Figure 1 shows such a device 1 (orange) that detects the voice V1 of a user H (purple), processes it, and outputs a responsive voice V2:

FIG.1

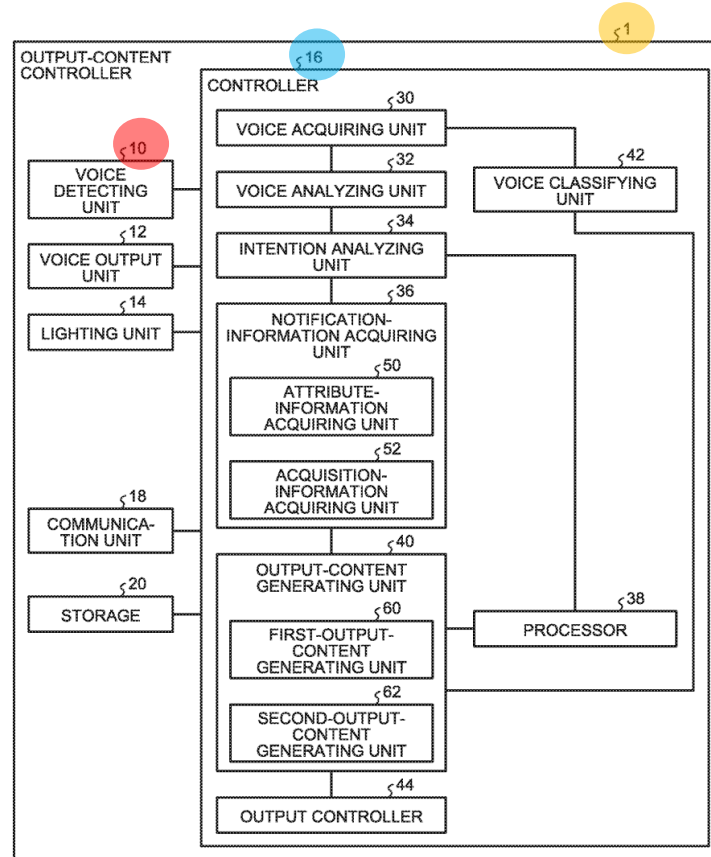


(*Id.*, Fig. 1, 2:64-3:4.)¹

45. The device 1 includes various components, such as a voice detecting unit 10 (red) and a controller 16 (blue):

¹ Figures and Tables in this Declaration may be colored or otherwise annotated for clarity.

FIG.2



(Ex. 1001 ('675 patent), Fig. 2, 3:17-22.) The controller includes several “units” for acquiring the voice, analyzing it, processing the request, and generating a response. (*Id.*, 3:55-67.) For example, the intention analyzing unit analyzes the user’s speech (e.g., “[T]oday’s schedule is[?]”) and the notification-information acquiring unit obtains the requested information (e.g., meeting, Mr. Yoshida, 15 o’clock, and Tokyo building). (*Id.*, 3:27-28, 4:49-57, 8:30-44.)

46. The device may include a “proximity sensor” for calculating the distance to the user. (Ex. 1001 ('675 patent), 10:32-38.) This generic sensor, described in a single sentence, purportedly allows the device to classify a voice as either a first

voice or a second voice. (*Id.*) When the voice is classified as a first voice (e.g., beyond a threshold distance), the device may acquire information to generate a first sentence (e.g., “Meeting with Mr. Yoshida from 15 o’clock at Tokyo building”), replace a word (e.g., Mr. Yoshida) in that information with another word (e.g., Yoshi), and output the generated sentence with the replacement word. (*Id.*, 10:32-38, 10:42-46, 8:55-63, 11:53-12:55, 14:40-43, claims 1, 6-7.) When the voice is classified as the second voice (e.g., within a threshold distance), the device may generate and output a second sentence that uses the original word. (*Id.*; *id.*, 11:5-52.)

B. Priority Date

47. I understand that the ’675 patent issued on February 8, 2022, from U.S. Patent Application 16/295,034 (“the ’034 application”), filed March 7, 2019. I also understand that the ’034 application claims priority to Japanese Application No. 2018-044598, filed on March 12, 2018. Thus, I understand that the ’675 patent’s earliest possible priority date is March 12, 2018.

VI. CLAIMS 1-7 OF THE ’675 PATENT WOULD HAVE BEEN OBVIOUS.

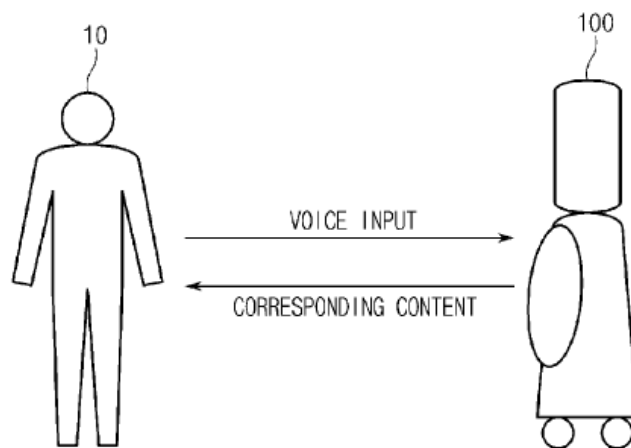
A. Claims 1-7 Would Have Been Obvious in View of Shin and Aoyama.

48. For at least the reasons I discuss below, Shin and Aoyama render obvious claims 1-7.

1. Claim 1

a. 1[pre]: Output-Content Control Device

49. The preamble of claim 1 recites an “output-content control device.” Shin discloses an electronic device that receives a user’s speech input through a microphone, “generate[s] content corresponding to a result of analyzing the voice input,” and “provide[s] the generated content as sound through an embedded audio output module (e.g., a speaker).” (Ex. 1003 (Shin) ¶[0037]; *see also id.*, Abstract, ¶¶[0003], [0008]-[0011], [0038]-[0039], [0048]-[0060], [0069]-[0072], claims 1, 15, 20.) For example, if a user says, “Let me know what time it is now,” Shin’s device may respond with, “The current time is nine ten AM.” (*Id.* ¶[0037].) The device may be a smartphone, PC, or home appliance. (*Id.* ¶¶[0031]-[0032].) Figure 1A shows an example of device 100:



(*Id.*, Fig. 1A.) As I discuss below, the device controls the output content. (*See* ¶¶85-108, below.) Thus, Shin discloses an “output-content control device.”

b. 1[a]: Voice Acquiring Unit

50. Claim element 1[a] recites “a voice acquiring unit configured to acquire a voice spoken by a user.” Shin discloses this claim element.

51. The '675 patent states that a voice acquiring unit may be part of a controller, which performs processes by reading software/program stored in a storage. (Ex. 1001 ('675 patent), 3:55-67, Fig. 2.) This controller can be a central processing unit (CPU). (*Id.*) Thus, according to the '675 patent, the “voice acquiring unit” refers to a CPU configured to perform the claimed function, namely, to acquire a voice spoken by a user.

52. Shin’s device 101 (orange) includes a processor 120 (blue) and an audio input module 151 (red):

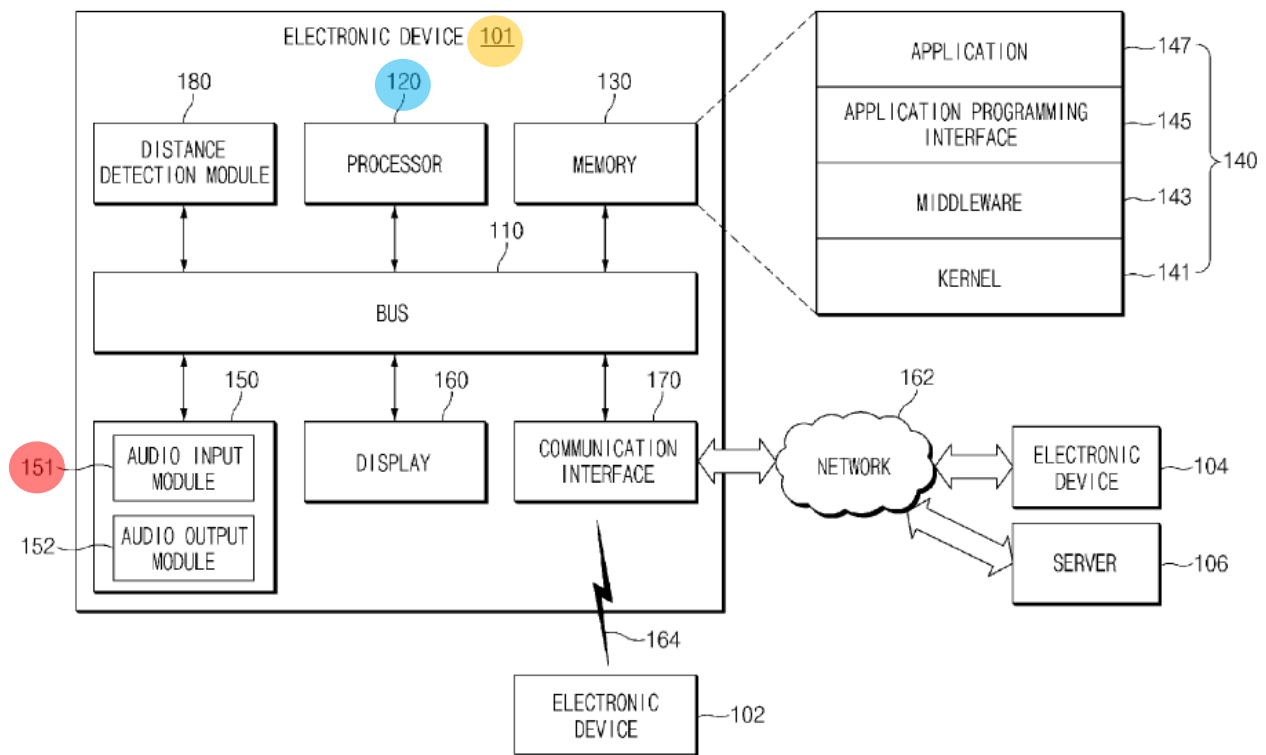


FIG. 3

(Ex. 1003 (Shin), Fig. 3, ¶[0045].) Processor 120, which may be a CPU (*id.* ¶[0029]), “analyze[s] a voice input received through [the] audio input module 151.” (*Id.* ¶[0048]; *see also id.*, Abstract, ¶¶[0003]-[0011], [0037]-[0039], [0041], [0060]-[0061], claims 1, 15, 20.) The audio input module 151 can be “implemented with a microphone and the like” to “obtain a user’s speech as a voice input.” (*Id.* ¶[0062]; *see also id.* ¶¶[0037], [0155], [0158].)

53. Thus, Shin discloses a voice acquiring unit (e.g., audio input module and/or portion of processor that receives speech from the audio input module) configured to acquire a voice spoken by a user.

c. 1[b]: Voice Classifying Unit

54. Claim element 1[b] recites “a voice classifying unit configured to [i] calculate a distance between the user and the output-content control device by a proximity sensor to [ii] classify the voice into either a first voice or a second voice based on the calculated distance.” Shin discloses or renders obvious this claim element.

55. The ’675 patent explains that, like the voice acquiring unit, a voice classifying unit may be part of a controller that performs processes by reading software/program stored in a storage. (Ex. 1001 (’675 patent), 3:55-67, Fig. 2.) Thus, the “voice classifying unit” refers to a CPU configured to perform the claimed function, namely, to calculate a user-to-device distance and classify a user’s voice based on the calculated distance.

i. 1[b][i]: Calculates Distance

56. Shin discloses that its processor “compute[s] the distance between the user and the electronic device[.]” (Ex. 1003 (Shin) ¶[0066].) This may be performed based on image data obtained from a distance detection module. (*Id.*; *see also id.* ¶¶[0053] (“the processor 120 may determine a distance between the user and the electronic device 101 based on ... the distance computed, calculated, or measured by the distance detection module 180”), [0075], [0077], Fig. 5A.) Shin’s distance detection module 180 may include a proximity sensor, such as “a depth camera like

a time-of-flight (TOF) camera, a stereo camera computing depth information using triangulation, a charge coupled device (CCD) camera computing a distance through an image processing, or the like” and/or “various sensors,” such as “an infra-red sensor, an RF sensor, an ultrasonic sensor, and the like.” (*Id.* ¶[0066]; *see also id.* ¶[0042].) Shin’s Figure 3 shows the electronic device 101 comprising processor 120 (blue) and distance detection module 180 (green):

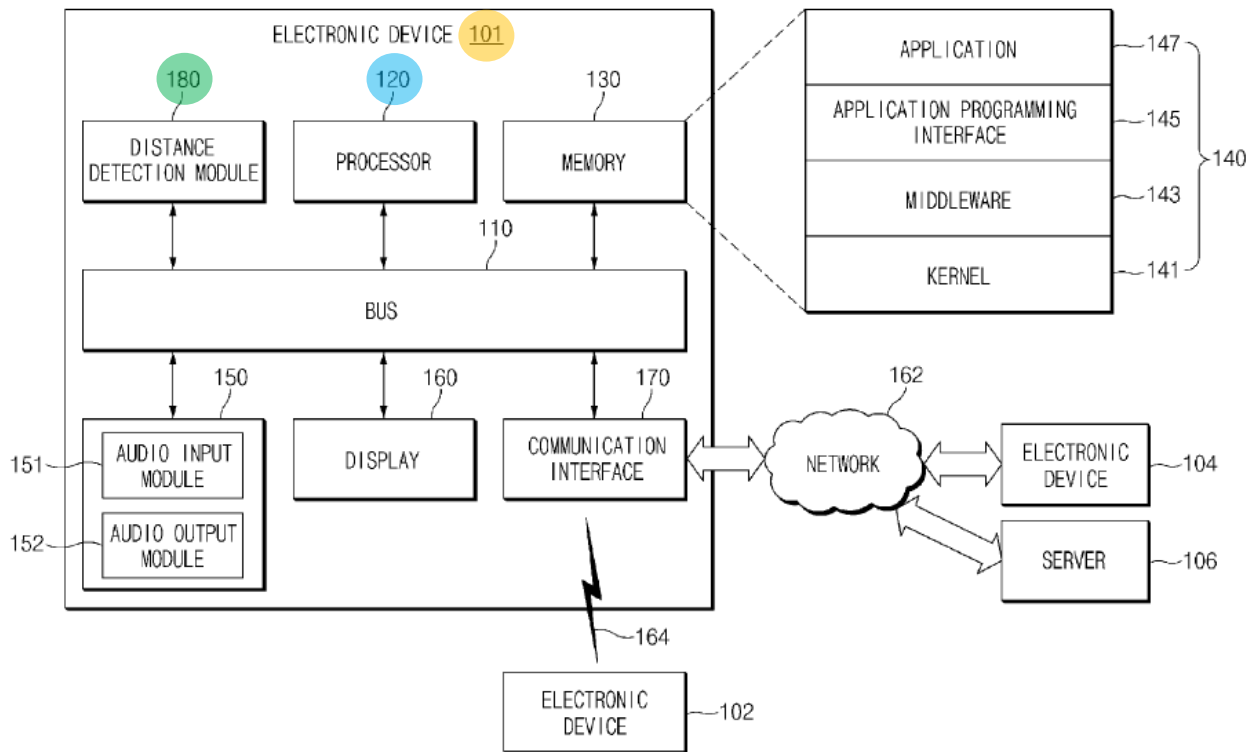


FIG. 3

(*Id.*, Fig. 3, ¶[0045]; *id.* ¶¶[0053]-[0054], [0077], [0096], claims 5, 9.) I note that, while Figure 3 shows “electronic device 101,” Shin discloses that its electronic device 100 may be implemented with the modules of electronic device 101. (*Id.* ¶¶[0041], [0043].)

57. Shin also expressly discloses that the electronic device 101 may use “a proximity sensor.” (Ex. 1003 (Shin) ¶[0153], Fig. 9 (“proximity sensor” 940G in “sensor module” 940); *see also id.* ¶[0143] (device 901 may be included in devices 100, 101).)

58. Thus, Shin discloses a voice classifying unit (processor) configured to calculate a distance between the user and the output-content control device by a proximity sensor.

ii. 1[b][iii]: Classifies Voice Based on Distance

59. The '675 patent explains that, when a proximity sensor is used for calculating user-to-device distance, the distance can be used as a “feature value” to perform the classification as the first or second voice. (Ex. 1001 ('675 patent), 10:32-38.) Specifically, the voice classifying unit “sets a threshold of the feature value, and classifies the voice” as the first or second voice “based on whether the feature value exceeds the threshold.” (*Id.*, 10:42-46.) Shin discloses the same thing.

60. Shin discloses that its processor analyzes the voice to classify it as either a first or second voice based on the user-to-device distance. Shin’s processor executes a speech recognition application to process the speech input and generates corresponding content for output. (Ex. 1003 (Shin) ¶¶[0047], [0060]; *see also id.* ¶¶[0037], [0039].) The processor determines the “output scheme,” which includes the content, “based on the distance between the user and the electronic device 101[.]”

(*Id.* ¶¶[0078], [0086]; *see also id.* ¶¶[0053], [0077], claim 17.) To do so, Shin’s processor classifies the voice based on the distance.

61. An example is shown in Shin’s Table 3, which discloses providing a different amount of information based on the user-to-device distance:

TABLE 3

Distance (D) between user and electronic device	Output amount of information of content
$D < 1\text{ m}$	Amount of information of original content
$1\text{ m} \leq D < 2\text{ m}$	Content of first abbreviation level
$2\text{ m} \leq D < 3\text{ m}$	Content of second abbreviation level
$3\text{ m} \leq D < 4\text{ m}$	Content of third abbreviation level
$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

(Ex. 1003 (Shin), Table 3 (consolidated).) For example, Shin discloses classifying the voice as a “first voice” when the distance is between 1 to 2 meters, and a “second voice” when the distance is less than 1 meter, as shown in Table 3:

TABLE 3

	Distance (D) between user and electronic device	Output amount of information of content
Second Voice	$D < 1\text{ m}$	Amount of information of original content
First Voice	$1\text{ m} \leq D < 2\text{ m}$	Content of first abbreviation level
	$2\text{ m} \leq D < 3\text{ m}$	Content of second abbreviation level
	$3\text{ m} \leq D < 4\text{ m}$	Content of third abbreviation level
	$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

(*Id.*; *see also id.* ¶¶[0051] (content may be “classified dichotomously”), [0086]-[0087].) Any of the distance ranges beyond 1 meter would satisfy the claimed “first

voice.” For example, a voice between 3 and 4 meters away could be the recited “first voice.” Alternatively, all distances greater than 1 meter could collectively be a first voice (e.g., a voice greater than 1 meter away).

62. Thus, Shin discloses that the voice classifying unit (processor) is configured to classify the voice as either a first voice (e.g., between 1 and 2 meters) or a second voice (e.g., closer than 1 meter) based on the calculated distance.

63. Even if claim 1 were to require the voice to be classified into no more than two classifications, Shin would render this obvious. A person of ordinary skill in the art would have been motivated to implement Shin’s amount-of-information table (Table 3) with only two distance ranges for several reasons.

64. First, Shin suggests it. Shin discloses dividing content into two classifications. (Ex. 1003 (Shin) ¶[0051].) Based on that disclosure, a person of ordinary skill would have been motivated to also use two distance ranges, one for each classification.

65. Second, Shin discloses that its “embodiments ... are not limited to the example[] of ... Table 3.” (Ex. 1003 (Shin) ¶[0090].) This flexibility would have motivated a person of ordinary skill to implement a simpler, two-classification system—such as “near” and “far”—sufficient for applications requiring only basic distinctions to adjust content output based on user distance.

66. Third, simplifying complex systems into fewer categories is a routine and well-established approach to reduce computational load and improve efficiency. Binary classification (classifying into two categories) is a standard approach in many fields, including voice-controlled devices (*see, e.g.*, Ex. 1004 (Shimomura), Figs. 7, 12), to simplify user experience. Many consumer devices adopt binary distinctions, such as “low” and “high” or “near” and “far.” A person of ordinary skill would have recognized binary classification as an efficient, user-friendly solution for Shin’s needs. For example, many systems developed in my own work rely on binary classification such as detecting the presence or absence of speech (*e.g.*, Vuong et al., 2021, 2023) or determining whether the source of detected speech is actually a person in the room or a recording (*e.g.*, Lindsey et al., 2023).

67. Fourth, such a modification reflects merely applying a known technique (binary classification) to achieve predictable results (a streamlined, efficient, and straightforward user interface with content output adjusted based on either “near” or “far” proximity classifications).

68. Fifth, such a modification reflects merely applying a known technique (binary classification) to a known device (Shin’s) that is ready for improvement and yields predictable results (consistent and reliable adjustments to content detail based on user proximity threshold, enhancing user interaction).

69. A person of ordinary skill would have reasonably expected success in modifying Shin in this way. Doing so would have been trivial, involving configuring Shin’s processor with revised tables for two proximity-based categories and optimizing performance for specific uses.

70. Thus, Shin also renders obvious this claim element.

d. 1[c]-1[d]: Intention Analyzing Unit and Notification-Information Acquiring Unit

71. Claim element 1[c] recites “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.” Claim element 1[d] recites “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.” Shin discloses or renders obvious these claim elements.

72. The ’675 patent explains that, like the voice classifying unit, an intention analyzing unit and a notification-intention acquiring unit may be parts of a controller or CPU that perform processes by reading software/program stored in a storage. (Ex. 1001 (’675 patent), 3:55-67, Fig. 2.) The “intention analyzing unit” may use text data from a voice analyzing unit (e.g., “today’s schedule is”) and determine

what information the user wishes to acquire (e.g., details about Mr. Yamada’s schedule). (*Id.*, 4:15-5:57, 9:22-24.) Then, the “notification-intention acquiring unit” obtains notification information (e.g., meeting, Mr. Yoshida, 15 o’clock, Tokyo building) based on that desired information. (*Id.*, 5:4-6:23.) The ’675 patent admits that devices that performed these functions were known. (*Id.*, 1:20-27.)

73. Shin’s processor analyzes the voice acquired by the voice acquiring unit. (Ex. 1003 (Shin) ¶¶[0048], [0062].) Shin’s processor executes a speech recognition application to process the speech input (which requests content in which the user is interested) and generate corresponding content for output by analyzing the request. (*Id.* ¶¶[0048], [0058]-[0060], [0084], [0089]; *see also id.* ¶¶[0037], [0058], [0067], [0071], [0088], [0103], [0107], [0120], [0126]-[0127], [0141].)

74. Shin discloses analyzing the acquired voice “to detect intention information” and acquiring “notification information ... based on the intention information” in the same way as the ’675 patent. For example, Shin describes analyzing a user’s speech, such as “Let me know today’s weather,” to determine intention information (e.g., weather) and then acquiring notification information (e.g., sky conditions, air temperature, and rainfall) based on that intention information. (Ex. 1003 (Shin) ¶[0088].) In another example, Shin describes analyzing a user’s speech, such as “Let me know what time it is now,” to determine intention information (e.g., time)

and then acquiring notification information (e.g., current time) based on that intention information. (*Id.* ¶[0037]; *see also id.* ¶[0039].)

75. To obtain content, Shin’s device may store an application in memory that “analyz[es] a voice input[] and generat[es] and/or provid[es] the content corresponding to the voice input.” (Ex. 1003 (Shin) ¶[0060].) The device can also delegate some processing to another device (e.g., another electronic device or a server), which delivers the results. (*Id.* ¶[0067]; *see also id.* ¶¶[0064]-[0065], [0174]-[0176].) Accordingly, Shin discloses acquiring the information from memory or from another device.

76. Even if Shin had not expressly disclosed these claim elements, Shin inherently discloses them. Shin’s processor necessarily includes an “intention analyzing unit” and a “notification-information acquiring unit” because such units would have been required for Shin’s device to function as described. For example, to provide content such as weather or time information, Shin’s processor necessarily analyzes the voice input to determine the user’s intent (e.g., a request for weather or time) and necessarily acquires the corresponding content information (e.g., “cloudy” or “nine ten AM”). (Ex. 1003 (Shin) ¶¶[0037], [0088].) These functionalities are required for the processor to obtain and deliver the requested content as Shin describes. Thus, Shin expressly or inherently discloses these elements.

77. Even if Shin had not disclosed these claim elements, they would have been obvious over Shin. Shin’s processor analyzes voice input to identify content the user is interested in, then acquires and provides that content. A person of ordinary skill would have recognized that implementing an “intention analyzing unit” to extract the user’s intent and a “notification-information acquiring unit” to retrieve content based on that intent would have been a predictable and logical extension of Shin’s functionalities.

78. A person of ordinary skill in the art would have been motivated to structure Shin’s processor with an intention analyzing unit and a notification-information acquisition unit for greater efficiency. Such a modular design would have facilitated consistent intent extraction and reliable content retrieval, and would have improved organization of the device’s processes. It would also merely reflect the way voice assistants worked at the time. (*See* ¶33, above.) For example, Duta discloses that conventional voice analysis systems in 2014 included an “[i]ntent classifier” module that is used to “find[] the query intent (what the user asks for).” (Ex. 1015 (Duta), 431-432, Fig. 1.) Such systems also included a “dialog manager” module that is used to respond to the user, for example by asking for additional information or confirmation. (*Id.*, 432-433, Fig. 1.) Duta states that its description of conventional systems “generalizes all systems built in the past” and is illustrated in Figure 1:

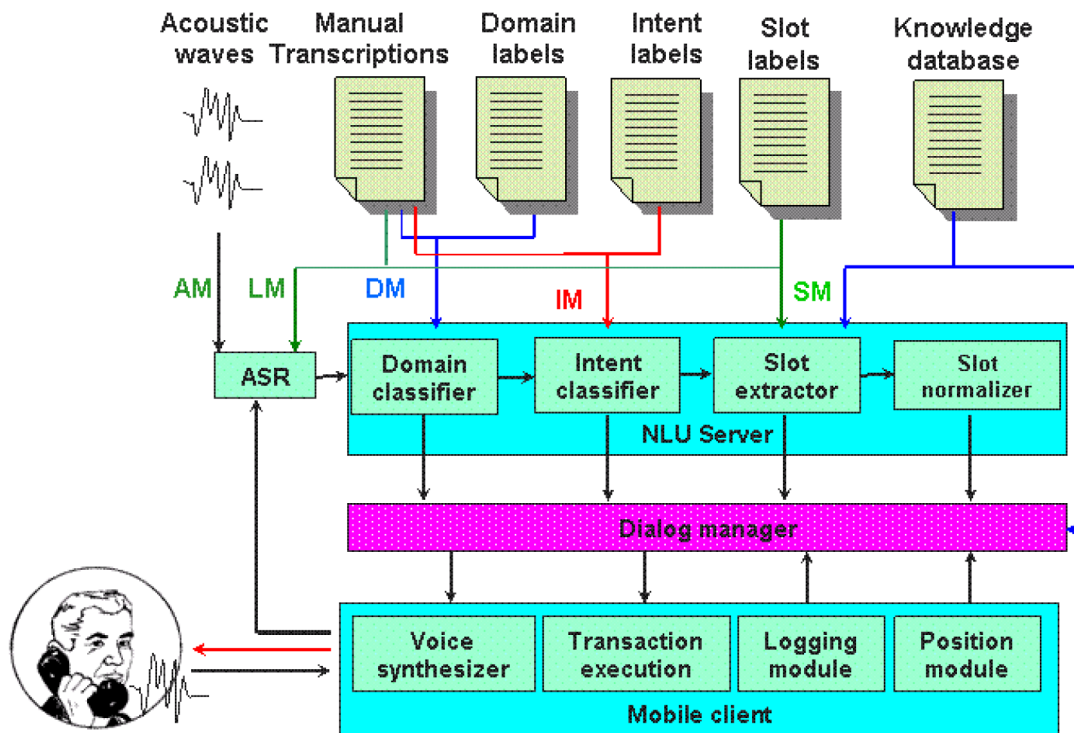


Figure 1. The architecture of a client-server voice-based transactional system including dialog.

(*Id.*, 432, Fig. 1.) Thus, this approach merely reflects known voice analysis techniques. (*Id.*, 431-433, Fig. 1.) This is further demonstrated by my discussion of Kristjansson below. (See ¶¶232-239, below.)

79. This approach would have been straightforward, as Shin's device already integrates speech analysis and content generation. Implementing these functionalities as processor units would have been an obvious design choice, enhancing modularity, maintainability, and troubleshooting. Using these units, Shin's device could improve response accuracy and handle diverse user requests with greater flexibility. Thus, Shin also renders obvious these claim elements.

80. Even if Shin did not disclose or render obvious these units, they are disclosed by and would have been obvious in view of Aoyama. Aoyama discloses a “meaning analysis unit” that uses natural language processing on the text of the user’s speech to determine the “meaning” of the input and specify a “task” to be executed. (Ex. 1010 (Aoyama) ¶¶[0072]-[0084], [0171]-[0173], [0230]-[0232], Figs. 3, 8, 13.) For example, for the input “Check my schedule for tomorrow at 1 pm,” Aoyama’s “meaning analysis unit” detects the intent information “schedule information.” (*Id.* ¶¶[0076]-[0079], [0170]-[0173].) This is the same as the intention analyzing unit in the ’675 patent. (Ex. 1001 (’675 patent), 4:49-57.) Thus, Aoyama discloses an intention analyzing unit (e.g., “meaning analysis unit”) configured to analyze the voice acquired by the voice acquiring unit to detect intention information (e.g., scheduling information) indicating what kind of information is wished to be acquired by the user.

81. Aoyama also discloses that a “response component generation unit” “executes the task indicated by the information acquired from the meaning analysis unit 114 ... and acquires information necessary to generate a response for the user[.]” (Ex. 1010 (Aoyama) ¶[0127]; *see also id.* ¶¶[0128]-[0130], [0174]-[0176], [0233], Figs. 3, 8, 13.) For example, in the scheduling example, Aoyama discloses that the response component generation unit “acquires a search result” from an application

that manages the schedule. (*Id.* ¶[0129].) Thus, Aoyama also discloses a notification-information acquiring unit configured to acquire notification information which includes content information (e.g., meeting title and participants) as a content information to be notified to the user based on the intention information (e.g., scheduling information).

82. A person of ordinary skill in the art would have been motivated to implement Aoyama’s “meaning analysis unit” and “response component generation unit” in Shin’s device for several reasons. First, Shin discloses a device that interprets and responds to voice inputs; a person of ordinary skill would have been motivated to look to other references, such as Aoyama, to understand how the speech input is processed, how tasks are executed, and how responses are generated.

83. Second, this combination would have reflected merely the addition of one known element (Aoyama’s natural language processing and information retrieval) to another known element (Shin’s voice processing and response device) to obtain predictable results (identifying a user’s intent and providing information responsive to such intent). Third, such a configuration reflects applying a known technique (Aoyama’s natural language processing) to improve a similar device and method (Shin’s content tailoring system) in the same way. Fourth, such a configuration applies a known technique (Aoyama’s natural language processing) to a

known device and method (Shin’s content tailoring system) that is ready for improvement and yields predictable results.

84. A person of ordinary skill in the art would have reasonably expected success in modifying Shin in this way because doing so would have been trivial, involving at most minor changes to Shin’s system to implement Aoyama’s units. Thus, Aoyama also renders obvious this claim element.

e. 1[e]: Output-Content Generating Unit

85. Claim element 1[e] recites “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.” Shin discloses this claim element.

86. In the ’675 patent, the “output-content generating unit” is the portion of the processor (CPU) that performs the recited function. (Ex. 1001 (’675 patent), 3:55-67, Fig. 2.) In one example, the “output-content generating unit” performs this function by generating the output sentence: “Meeting with Mr. Yoshida from 15 o’clock at Tokyo building.” (*Id.*, 8:55-63.) The patent admits that the claimed function was known. (*Id.*, 1:20-27.)

87. As I discussed above for claim elements 1[b]-1[d], Shin’s processor generates content for output to the user. (See ¶¶54-84, above.) In Shin’s time example, the content can be an output sentence including sentence data, such as: “The

current time is nine ten AM.” (Ex. 1003 (Shin) ¶[0037].) In Shin’s weather example, Shin’s processor performs the claimed function when it generates the output sentence: “The weather in Jul. 1, 2015 is (①) rainy after cloudy (②). The highest temperature is 28° C., and the lowest temperature is 18° C. (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).” (*Id.* ¶[0088].) These sentences may be converted into speech using text-to-speech technology and output audibly. (*Id.* ¶[0048]; *see also id.* ¶¶[0007], [0037], [0039], [0062], [0072], [0109], [0131], [0139], [0158].) Thus, Shin discloses an output-content generating unit (portion of processor) configured to generate an output sentence that is sentence data to be output to the user.

88. Shin’s output-content generating unit generates the output sentence based on notification information. As I discussed above, Shin’s processor analyzes the voice input to acquire notification information according to the user’s interests. (*See* ¶¶71-84, above.) This notification information is incorporated into Shin’s output sentences. In the weather example, the output sentences incorporate notification information like sky conditions, air temperature, and rainfall. (Ex. 1003 (Shin) ¶[0088].) In the time example, the output sentence incorporates notification information like the current time. (*Id.* ¶[0037].)

89. Thus, Shin discloses an output-content generating unit (processor) configured to generate an output sentence that is sentence data to be output to the user based on the notification information.

f. 1[f]: Generate a First Output Sentence

90. Claim element 1[f] recites that “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.” Shin and Aoyama render obvious this limitation.

91. As an example, the '675 patent's device may respond to a spoken request about what “today's schedule is.” (Ex. 1001 ('675 patent), 4:50-57.) The device may generate an output sentence in which “‘Mr. Yoshida’ is replaced with ‘Yoshi’” to produce the revised sentence: “Mr. Yamada's schedule on Mar. 20, 2020 is a meeting with Yoshi from 15 o'clock at Tokyo building.” (*Id.*, 14:40-43, 11:37-40; *see also id.*, Fig. 6 (“dinner” replaced with “meeting”).)

92. Shin's processor adjusts the amount of information provided to the user based on the user-to-device distance. (Ex. 1003 (Shin) ¶[0053]; *see also id.* ¶¶[0051], [0086]-[0090], claims 8, 17.) Shin discloses that, whenever the user-to-device “distance increases by 1 m, the electronic device 101 may decrease the output amount of information by making the abbreviation level of the original content

high.” (*Id.* ¶[0087]; *id.* ¶[0086] (“output amount of information of the content decreases as a distance between the user and the electronic device 101 increases”).)

Shin’s Table 3 shows this:

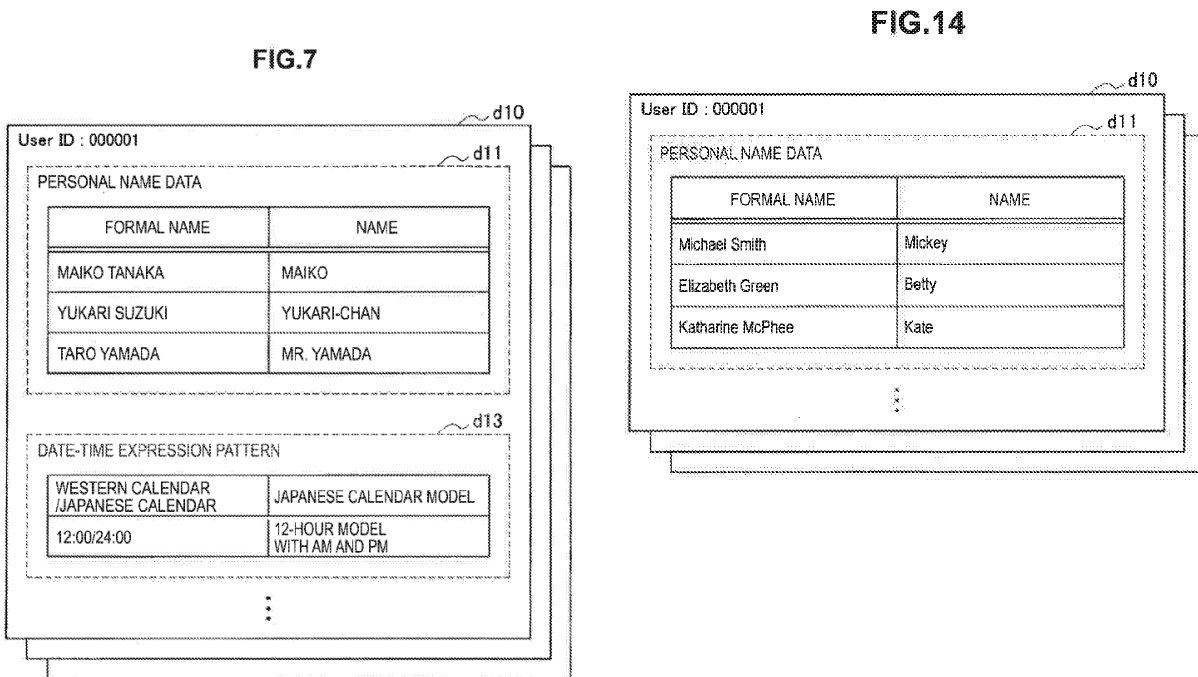
TABLE 3	
Distance (D) between user and electronic device	Output amount of information of content
$D < 1\text{ m}$	Amount of information of original content
$1\text{ m} \leq D < 2\text{ m}$	Content of first abbreviation level
$2\text{ m} \leq D < 3\text{ m}$	Content of second abbreviation level
$3\text{ m} \leq D < 4\text{ m}$	Content of third abbreviation level
$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

(*Id.*, Table 3.) As shown, when the user-to-device distance is between 1 and 2 meters, the voice falls into a first classification and the processor generates “[c]ontent of first abbreviation level,” which may be described more “simply” and abbreviated to include “a gist of the response” with less information. (*Id.* ¶¶[0051], [0087].) Indeed, at any distance beyond 1 meter, the device outputs abbreviated content. (*Id.*) A person of ordinary skill in the art would have readily understood Shin’s disclosure of abbreviating content to describe it more simply or provide the “gist” includes replacing words.

93. Even if Shin does not expressly disclose replacing words to generate abbreviated/simplified content, it would have been obvious over Aoyama. Aoyama discloses an electronic device (e.g., smartphone) with a processor that recognizes

and analyzes spoken words from a user, executes processing based on that input, and audibly presents the result as a sentence. (Ex. 1010 (Aoyama) ¶¶[0042], [0051], [0127], [0161], [0241]-[0243], [0253], Fig. 8.) Aoyama’s device generates customized output by replacing words based on the user or situation (e.g., selecting between formal and informal names). (*Id.* ¶¶[0120]-[0124], [0138]-[0140], [0177], [0218].)

94. Aoyama’s Figures 7 and 14 show “personal name data” tables, which relate formal names to informal names:



(Ex. 1010 (Aoyama), Figs. 7, 14; *id.* ¶¶[0120]-[0124], [0234]-[0236].)

95. Like the '675 patent, Aoyama discloses replacing names in response to a request about a user’s schedule. Aoyama can generate a response using a person’s

full name (e.g., Taro Yamada), honorific title and surname (e.g., Mr. Yamada), honorific title and full name (e.g., Mr. Taro Yamada), or first name only (e.g., Taro). (Ex. 1010 (Aoyama) ¶¶[0050], [0126]; *see also id.* ¶¶[0123]-[0125], [0133], [0138], [0140], [0225], [0235]-[0236], Fig. 7.) Aoyama also discloses replacing other words in an output sentence, such as the date and time. (*Id.* ¶[0140].) For example, Aoyama replaces the name “Taro Yamada” with the name “Mr. Yamada,” the date “10/1/2014” with “tomorrow,” and “13:00” with “1 pm.” (*Id.* ¶¶[0138]-[0140].) Thus, the device can generate an output sentence such as “You have a meeting with Mr. Yamada in room A tomorrow at 1pm,” instead of “You have a meeting with Taro Yamada in room A at 13:00 10/1/2014.” (*Id.*) Aoyama also discloses replacing words to refer to individuals by either their first name (e.g., Hirofumi or Yuri) or familial role (e.g., Papa or Mama). (*Id.* ¶¶[0221], [0217]; *see also id.* ¶[0154] (weather example where “clear” replaces “nice”).)

96. A person of ordinary skill in the art would have been motivated to configure Shin’s processor to adjust the information provided to the user based on the user-to-device distance by incorporating Aoyama’s word replacement feature for many reasons.

97. First, Aoyama’s word replacement feature furthers Shin’s primary goal: to simplify output for users at greater distances and make the information easier for

such users to grasp. For example, Aoyama discloses replacing full names with simple nicknames (replacing “Michael Smith” with “Micky” or Elizabeth Green with “Betty”) or familial names (replacing “Hirofumi” with “Papa”), and replacing complicated date and time formats with simpler terms, which would make the output easier for users to understand/grasp at greater distances. (Ex. 1010 (Aoyama) ¶¶[0221], [0235], [0138]-[0140].)

98. Second, Aoyama teaches the desirability of replacing words based on the user state (e.g., stopped, walking, running, laying, sitting, standing, their speech speed, age, sex, etc.). (Ex. 1010 (Aoyama) ¶¶[0097]-[0113].) Aoyama further teaches that other sensors can be used to acquire information about the user state. (*Id.* ¶[0114].) This disclosure would have motivated a person of ordinary skill to use a known proximity sensor, as in Shin, and replace words based on the user’s distance, which a person of ordinary skill would have understood to be another example of user state.

99. Third, it would have been obvious to combine Shin and Aoyama to address privacy concerns that were well-known in the art for years before the ’675 patent’s priority date. For example, Iyer discloses an electronic device that processes and audibly responds to speech commands in discreet and normal modes. (Ex. 1012 (Iyer), Abstract, 3:8-15.) Iyer notes that “a user may not want passersby to hear the audible output[,]” especially when the output includes personal information. (*Id.*,

3:19-27, 4:30-36.) Given known privacy concerns in voice systems like Shin's, a person of ordinary skill would have been motivated to incorporate Aoyama's word replacement feature into Shin's device to replace words containing personal information with neutral alternatives. Such privacy concerns were well known in the art. For example, Ha disclosed a device that outputs varying levels of information based on whether the user is in a public or private space, or whether there are other people nearby. (Ex. 1011 (Ha) ¶¶[0042]-[0045]; *see generally id.*)

100. Fourth, such a configuration merely represents the addition of one known element (Aoyama's word replacement) to another known element (Shin's distance-based content abbreviation/simplification) to obtain predictable results (providing content to users at greater distances that is abbreviated or simplified using word replacement). Fifth, such a configuration reflects applying a known technique (Aoyama's word replacement) to improve a similar device and method (Shin's content tailoring system) in the same way. Sixth, such a configuration applies a known technique (Aoyama's word replacement) to a known device and method (Shin's content tailoring system) that is ready for improvement and yields predictable results (a system that delivers content to users at greater distances that is tailored using word replacement to address privacy concerns).

101. Seventh, such a configuration would have been obvious to try. Shin's device presents only two design options for abbreviating or simplifying its output:

omitting words or replacing them. Given Shin’s disclosure of simplifying output to provide “core content” rather than “detailed content” for users farther away (Ex. 1003 (Shin) ¶[0089]), it would have been obvious to try a word-replacement approach, such as Aoyama’s.

102. A person of ordinary skill in the art would have reasonably expected success in incorporating Aoyama’s word replacement feature into Shin’s processor. This configuration would have been trivial, involving straightforward changes to enhance the processor’s approach to simplifying and abbreviating its output content. Furthermore, such an approach was well known in the art, further leading a person of ordinary skill to reasonably expect success. For example, Ha (filed by Samsung) discloses a device that replaces “team meeting” and “date” (described as “detailed information”) with “schedules” (described as “abstract information”) based on the user’s environment. (Ex. 1011 (Ha) ¶¶[0097]-[0098], Fig. 7; *see generally id.*)

103. Thus, Shin and Aoyama render this limitation obvious.

g. 1[g]: Generate a Second Output Sentence

104. Claim element 1[g] recites that the output-content generating unit is further configured 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.” Shin discloses this claim element.

105. The '675 patent provides an example of this limitation using the scheduling example I discussed above. (See ¶91, above.) In this example, the device generates the output sentence: “Mr. Yamada’s schedule on Mar. 20, 2020 is a meeting with Mr. Yoshida from 15 o’clock at Tokyo building.” (Ex. 1001 ('675 patent), 11:36-40.) In contrast with the example I discussed above, here, “Mr. Yoshida” is not replaced with “Yoshi.” (See ¶91, above.) In this example, the word “schedule” “indicate[s] the content of the intention information” while the words “Mr. Yamada,” “Mar. 20, 2020,” “Tokyo building,” “Mr. Yoshida,” “meeting,” and “15 o’clock” indicate the content of the “attribute content information” or “acquisition content information.” (Ex. 1001 ('675 patent), 11:26-37.)

106. Shin discloses that, when a voice is classified as a second voice (e.g., closer than 1 meter), the device generates a second output comprising all available information. (Ex. 1003 (Shin) ¶¶[0087] (when distance “is less than 1 m, an output amount of information of the content may be set highest” and provides “the most detailed original content corresponding to a voice input”), [0051].) The amount of information provided is shown in Table 3:

TABLE 3

	Distance (D) between user and electronic device	Output amount of information of content
Second Voice	$D < 1 \text{ m}$	Amount of information of original content
First Voice	$1 \text{ m} \leq D < 2 \text{ m}$	Content of first abbreviation level
	$2 \text{ m} \leq D < 3 \text{ m}$	Content of second abbreviation level
	$3 \text{ m} \leq D < 4 \text{ m}$	Content of third abbreviation level
	$D \geq 4 \text{ m}$	Content of fourth abbreviation level (abstract content)

(*Id.*, Table 3.) Thus, when classified as a second voice (e.g., closer than 1 meter), the processor outputs the content without abbreviation. (*Id.* ¶¶[0087]-[0088].)

107. Shin’s time example also illustrates Shin’s output of unabbreviated content with the output sentence: “The current time is nine ten AM.” (*Compare* Ex. 1003 (Shin) ¶[0037] *with id.* ¶[0039] (shortened reply “nine ten”).) In this example, the word “time” conveys all intention information derived from the user statement, “Let me know what time it is now.” The words “nine ten” represent all content information determined for that user statement.

108. Accordingly, for at least the reasons I discuss above, Shin and Aoyama render obvious claim 1.

2. Claim 2

109. Claim 2 depends from claim 1 and recites two additional elements, which Shin and Aoyama render obvious.

a. 2[a]: Notification Information

110. Claim element 2[a] recites that “the notification information includes type information indicating a type of the content information.” Aoyama discloses this claim element.

111. The ’675 patent describes “type information” as information classifying the user’s desired content. (Ex. 1001 (’675 patent), 7:15-19.) For instance, in the ’675 patent’s scheduling example, the notification information identifies who the user is scheduled to meet with type information (“person”) and content information (“Mr. Yoshida”). (*Id.*, 8:32-44, Fig. 5.)

112. As I discussed above, Aoyama’s device references replacement words stored in tables, such as one titled “personal name data,” shown in Figure 7:

FIG.7

User ID : 000001

FORMAL NAME	NAME
MAIKO TANAKA	MAIKO
YUKARI SUZUKI	YUKARI-CHAN
TARO YAMADA	MR. YAMADA

WESTERN CALENDAR / JAPANESE CALENDAR	JAPANESE CALENDAR MODEL
12:00/24:00	12-HOUR MODEL WITH AM AND PM
⋮	

(Ex. 1010 (Aoyama), Fig. 7; *see* ¶¶90-103, above.) Aoyama discloses that its system includes a “response generation unit 17” that generates the device’s response to the user. (Ex. 1010 (Aoyama) ¶[0124].) Another unit, “the response parameter generation unit 16 ... receive[s] notification of the name corresponding to ‘Taro Yamada’ from the response generation unit 17.” (*Id.*) Thus, Aoyama’s system internally transmits the content information (e.g., “Taro Yamada”) along with type information (e.g., “name”). (*Id.*) The response parameter generation unit then looks up the name in the appropriate table, which it determines using the type information. (*Id.*) It then “notifies the response generation unit 17 of information indicating the ‘name’ the recognized [sic] ‘Mr. Yamada’” so that “the response generation unit 17 can generate a response” using the type and content information. (*Id.*) Thus, Aoyama discloses that the notification information includes type information (e.g., “name”) indicating a type of the content information (e.g., “Taro Yamada”).

113. Similarly, the titles of Aoyama’s tables are type information because they correspond to the types of content information the system looks up and replaces. For example, “name” in the title “personal name data” is type information for content information such as “Taro Yamada” and “Mr. Yamada” in Aoyama, just as “person” is type information for content information such as “Yoshi” and “Mr. Yoshida” in the ’675 patent. (Ex. 1001 (’675 patent), 14:41-43, 11:37-40.)

b. 2[b]: Output-Content Generating Unit

114. Claim element 2[b] recites that “the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be 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.” I have been instructed to assume that “placed” is intended to mean “replaced.”

115. The '675 patent discloses that, after generating the sentence “Mr. Yamada’s schedule on Mar. 20, 2020 is a meeting with Mr. Yoshida from 15 o’clock at Tokyo building,” the device selects the words “Mr. Yoshida” to be replaced with the word “Yoshi.” (Ex. 1001 ('675 patent), 14:40-43, 11:37-40.) This selection is based on “Mr. Yoshida” appearing in the content information and the type information “person” associated with “Mr. Yoshida.” (*Id.*, 14:27-43, 12:50-54.) Aoyama also discloses this.

116. As I discussed above, Aoyama discloses replacing names, including replacing “the response component indicating the personal name ‘Taro Yamada’ with the name ‘Mr. Yamada’ in accordance with the speech style of the user.” (See ¶¶90-103, above.) Aoyama discloses that the “‘speech style of the user’ includes personal name data,” which “is data for managing the speech style regarding calling of a personal name and is stored by associating a ‘formal name’ and a ‘name’ corresponding

to the ‘formal name’ with each person.” (Ex. 1010 (Aoyama) ¶[0121].) Thus, Aoyama discloses selecting a word (e.g., “Taro”) to be replaced with another word (e.g., “Mr.”) based on a word (e.g., “Taro Yamada”) included in the content information and the type information (e.g., “name”) associated with that content information. (*Id.*) Thus, Aoyama discloses this limitation.

117. Accordingly, for at least the reasons I discuss above, Shin and Aoyama render obvious claim 2.

3. Claim 3

118. Claim 3 depends from claim 2 and further recites that “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.” Aoyama discloses this limitation.

119. The ’675 patent discloses an example where relationship information is read from “a contact list including names of persons” (e.g., a table) in which “different names (nickname or the like) are designated to the names of the persons.” (Ex. 1001 (’675 patent), 14:23-27, 14:12-13.) In this example, the device selects the words “Mr. Yoshida” in the notification information to be replaced with the word

“Yoshi,” since “Mr. Yoshida” in the notification information matches “Mr. Yoshida” in the relationship information and “Mr. Yoshida” is associated with “Yoshi” in the relationship information. (*Id.*, 14:36-43.)

120. Aoyama’s device references replacement words from tables when generating an output sentence. (See ¶¶90-103, 114-117, above.) For example, Aoyama’s device reads the table titled “personal name data,” which “associat[es] a ‘formal name’ and a ‘name’ corresponding to the ‘formal name’ with each person,” as shown in Figure 7:

FIG.7

User ID : 000001

PERSONAL NAME DATA

FORMAL NAME	NAME
MAIKO TANAKA	MAIKO
YUKARI SUZUKI	YUKARI-CHAN
TARO YAMADA	MR. YAMADA

DATE-TIME EXPRESSION PATTERN

WESTERN CALENDAR / JAPANESE CALENDAR	JAPANESE CALENDAR MODEL
12:00/24:00	12-HOUR MODEL WITH AM AND PM

⋮

(Ex. 1010 (Aoyama) ¶[0121], Fig. 7.) These tables, including the associations listed therein, are relationship information. (See Ex. 1001 ('675 patent), 14:27-36.) For example, the table shows that “Taro Yamada” is related to “Mr. Yamada.” The table is stored in advance of generating the output sentence, so the words in the table are predetermined. (Ex. 1010 (Aoyama) ¶¶[0119], [0124]; *see also id.* ¶¶[0121]-[0122], [0198].) Thus, Aoyama’s device is 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.

121. Aoyama’s device also replaces a word included in the output sentence’s notification information with the second word when it matches the first predetermined word. As discussed above, Aoyama’s device replaces words in the output sentence to adjust name presentation, such as by “replac[ing] the response component indicating the personal name ‘Taro Yamada’ with the name ‘Mr. Yamada’ in accordance with the speech style of the user.” (See ¶¶90-103, above.) This replacement enables the device to generate content like “You have a meeting with Mr. Yamada in room A,” where “Mr. Yamada” replaces “Taro Yamada.” (Ex. 1010 (Aoyama) ¶[0138].) Thus, Aoyama’s device is configured 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.

122. Accordingly, for at least the reasons I discuss above, Shin and Aoyama render obvious claim 3.

4. Claim 4

123. Claim 4 depends from claim 3 and recites two additional elements, which Aoyama discloses.

a. 4[a]: Relationship Information

124. Claim element 4[a] recites that “the relationship information further includes information of type information indicating a type of the first word.” Aoyama discloses this limitation.

125. The ’675 patent’s contact list example (*see* ¶119, above) describes that relationship information includes type information “person,” which indicates the type of the words “Mr. Yoshida.” (Ex. 1001 (’675 patent), 14:30-36.)

126. Aoyama’s replacement word tables include titles that provide type information. For example, the table in Aoyama titled “personal name data” “associate[s] a ‘formal name’ and a ‘name’ corresponding to the ‘formal name’ with each person.” (Ex. 1010 (Aoyama) ¶[0121], Fig. 7; *see also id.* ¶[0235], Fig. 14.) The phrase “personal name” in the title “personal name data” indicates the type of the words “Taro Yamada” and “Mr. Yamada” listed in the table. Thus, Aoyama discloses this limitation.

b. 4[b]: Output-Content Generating Unit

127. Claim element 4[b] recites that “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.” Aoyama discloses this limitation.

128. In the '675 patent's example, the device selects the words “Mr. Yoshida” in the notification information to be replaced with “Yoshi.” (Ex. 1001 ('675 patent), 14:40-43.) This selection is based on two conditions: (i) the words “Mr. Yoshida” in the notification information matching the words “Mr. Yoshida” in the relationship information, and (ii) the type information “person” associated with “Mr. Yoshida” in the notification information matching the type information of “Mr. Yoshida” in the relationship information. (*Id.*, 14:36-39.)

129. As I discussed above, Aoyama's device selects a word (e.g., “Taro Yamada”) in the notification information to be replaced with another word (e.g., “Mr. Yamada”) based on the relationship information in Aoyama's tables. (*See* ¶¶114-117, above.) Aoyama's device determines this replacement by determining that the initial response includes the words “Taro Yamada,” which have a type “name.” (Ex. 1010 (Aoyama) ¶[0124].) The device then consults the “personal

name data” table, which includes data regarding words of type “name” that is personal to the user. (*Id.*; *see id.* ¶[0121].) Thus, Aoyama requires that the type information (e.g., “name”) for the word in the notification information (e.g., “Taro Yamada”) matches the type information for the word in the relationship information (e.g., “name”), because the type information determines which table the device will consult. (*Id.*) Then, the device finds the entry in the personal name data table for “Taro Yamada,” thereby also requiring that the word in the notification information (e.g., “Taro Yamada”) matches the word in the relationship information (e.g., “Taro Yamada”). (*Id.*) Thus, Aoyama discloses this limitation. (*Id.*; *see also id.* ¶¶[0119], [0138], Figs. 7, 14.)

130. Accordingly, for at least the reasons I discuss above, Shin and Aoyama render obvious claim 4.

5. Claim 5

131. Claim 5 depends from claim 1 and further recites that “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.” I note that the phrase “the predetermined voice” lacks an antecedent basis; I have been asked to assume that “the predetermined voice” refers to either “the first voice” or “the second voice” as recited in claim 1. Shin discloses or renders obvious this limitation.

132. Shin's device can determine a user's speech volume and adjust its output based, at least in part, on that volume. (Ex. 1003 (Shin) ¶[0071]; *see also id.* ¶¶[0037]-[0039], [0049]-[0050], [0053]-[0054], [0076]-[0078], [0095], [0137], Abstract, claims 1-9, 15-17, 20.) Shin discloses that its device detects a range of speech volumes and rates. (*Id.* ¶¶[0037]-[0038].) Accordingly, a person of ordinary skill in the art would have understood Shin's disclosure to encompass receiving and processing as wide a range of speech as possible, including whispered speech.

133. Even if Shin's device did not explicitly support processing whispered speech, a person of ordinary skill would have been motivated for several reasons to configure the device to receive and process such inputs.

134. First, the prior art suggests it. For example, Iyer teaches that the privacy concerns I discussed above (*see* ¶99, above) may be addressed in part by processing whispered speech. (Ex. 1012 (Iyer), 3:61-64 (“[W]hen in the discreet mode of operation, the user may whisper voice commands to the microphone.”).) This would have motivated a person of ordinary skill to have Shin's device process whispered speech to address privacy concerns. Furthermore, Shin's disclosure of processing speech at a wide range of volumes would have motivated a person of ordinary skill to process whispers, which a person of ordinary skill would have recognized are often quiet.

135. Second, such a configuration would represent merely the simple addition of a known element (volume detection including whisper) to another known element (Shin's volume detection for a range of speech volumes) to obtain predictable results (classifying speech input across wide range of speech). Third, such a configuration reflects merely a combination of known elements (detecting a wide range of speech, including whispers) to achieve an expected result (classifying speech input across a wide range of volumes). Fourth, such a configuration applies a known technique (detecting a wide range of speech, including whispers) to a known device and method (Shin's) that is ready for improvement and yields predictable results (classifying speech input across wide range of volumes).

136. Fifth, such a configuration would have been obvious to try. As discussed above, Shin's device, which determines a user's speech volume and adjusts its output accordingly (*e.g.*, Ex. 1003 (Shin) ¶[0071]), demonstrates the foundational capability to process speech inputs of varying intensities and types. (*Id.* ¶¶[0037]-[0038].) Extending this functionality to include whispers would have been an obvious and predictable variation.

137. A person of ordinary skill would have reasonably expected success in modifying Shin in this way. Shin's device already detects and responds to variations in speech volume. (Ex. 1003 (Shin) ¶¶[0037]-[0038], [0071].) Extending its capabilities to handle inputs like whispers would involve straightforward adjustments (if

any). Furthermore, such functionality was known in the art, further leading a person of ordinary skill to reasonably expect success. (*E.g.*, Ex. 1012 (Iyer), 3:61-64 (“[W]hen in the discreet mode of operation, the user may whisper voice commands to the microphone.”); *see generally* Ex. 1016 (Schuster); *see* ¶¶225-231, below.)

138. Accordingly, for at least the reasons I discuss above, Shin and Aoyama render obvious claim 5.

6. Claims 6 and 7

139. The preamble of claim 6 recites “[a]n output-content control method.” Shin discloses both electronic devices and methods of operating the electronic devices. (Ex. 1003 (Shin) ¶[0181]; *see also id.* ¶¶[0003], [0008]-[0011], [0016]-[0021], [0043], [0047], [0092]-[0093], Title, Abstract, claims 15-19.) Furthermore, the steps of claim 6 generally correspond to elements of claim 1 and are therefore rendered obvious by Shin and Aoyama for the reasons I discuss above.

140. The preamble of claim 7 recites “[a] non-transitory storage medium that stores an output-content control program that causes a computer to execute” a series of steps, which are the same as the steps recited in claim 6. Thus, the steps recited in claim 7 are rendered obvious by Shin and Aoyama for the reasons I discuss above. And, Shin discloses a non-transitory storage medium. Specifically, Shin discloses

electronic devices implemented by instructions stored in a computer-readable storage media. (Ex. 1003 (Shin) ¶[0181]; *see also id.* ¶¶[0011], [0179]-[0180], [0182], claim 20.)

141. I have compared the elements in claims 6 and 7 to the elements in claim 1. As shown, other than their preambles, claims 6 and 7 are identical and merely recite the functions performed by the various units recited in claim 1. The table below demonstrates how the elements correspond. As I discuss above for claim 1, Shin and Aoyama disclose all the steps recited in claims 6 and 7 and thus render obvious claims 6 and 7. (*See* ¶¶49-108, above.)

	Claim 4	Claim 5		Claim 1
6[pre]/7[pre]	An output-content control method comprising:	A non-transitory storage medium that stores an output-content control program that causes a computer to execute:	1[pre]	An output-content control device, comprising:
6[a]/7[a]	acquiring a voice spoken by a user;		1[a]	a voice acquiring unit configured to <i>acquire a voice spoken by a user</i> ;
6[b]/7[b]	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;		1[b]	a voice classifying unit configured to <i>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</i>

			<i>voice based on the calculated distance;</i>
6[c]/7[c]	analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;	1[c]	an intention analyzing unit configured to <i>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;</i>
6[d]/7[d]	acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and	1[d]	a notification-information acquiring unit configured to <i>acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and</i>
		1[e]	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,
6[e]/7[e]	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	1[f]	wherein the output-content generating unit is further configured to <i>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</i>

			<i>is replaced with another word, and</i>
6[f]/7[f]	generating, when the voice is not determined to be the second voice, a second output sentence which includes all of the intention information and the content information.	1[g]	<i>to generate, when the voice is not determined to be the second voice, a second output sentence which includes all of the intention information and the content information.</i>

B. Claims 1-2 and 5-7 Would Have Been Obvious in View of Shin and Iwase.

142. As I have discussed above, Shin discloses certain limitations of claims 1-7 and, with Aoyama, renders those claims obvious. (See ¶¶48-141, above.) The disclosure of Shin I discussed above also renders claims 1-2 and 5-7 obvious in view of Iwase. To avoid repetition, I will address below only the claims and limitations for which Iwase is relevant.

1. Claim 1

143. Iwase discloses, or at least suggests, the elements of claim 1 that recite replacing one word in the output content with another. For example, claim element 1[f] recites that “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.”

144. Iwase discloses an electronic device (e.g., smartphone) with a processor that audibly presents notification information to users. (Ex. 1017 (Iwase), 3:26-29, 5:2-7, 5:10-13, 5:63-6:3, 7:9-11, 13:47-49, Abstract.) To enhance comprehension, the device adjusts the notification based on user attributes, such as age or word preferences. (*Id.*, 1:54-2:18, 4:17-29, 6:32-35, 12:26-42, 15:6-40.) This adjusting may involve word replacement, such as replacing “I recommend” in the phrase “I recommend bringing an umbrella” with “You should,” yielding: “You should take an umbrella.” (*Id.*, 15:6-14.)

145. A person of ordinary skill in the art would have been motivated to configure Shin’s processor to adjust the information provided to the user based on the user-to-device distance by incorporating Iwase’s word replacement feature, for example to replace harder-to-understand words with simpler ones when the user is far away, for several reasons.

146. First, Iwase suggests doing so. Iwase discloses that its word replacement system allows a user “to grasp notification contents more effectively.” (Ex. 1017 (Iwase), 2:11-14, 15:24-28, 21:66-22:8.) A person of ordinary skill would have been motivated to incorporate Iwase’s word replacement feature into Shin’s device to gain the benefits disclosed by Iwase. (*Id.*)

147. Second, such a configuration represents the simple addition of one known element (Iwase's word replacement) to another known element (Shin's distance-based output adjustment) to obtain predictable results (easier-to-understand output to users at greater distances). Third, such a configuration reflects applying a known technique (Iwase's word replacement) to improve a similar device and method (Shin's content output system) in the same way (to make the output easier to grasp). Fourth, such a configuration also applies a known technique (Iwase's word replacement) to a known device and method (Shin's content abbreviation system) that is ready for improvement and yields predictable results (a system that delivers content to users at greater distances that is abbreviated using word replacement).

148. Fifth, such a configuration would have been obvious to try. Shin's device presents only two design options for simplifying and abbreviating its output content: removing words or replacing them. Given Shin's disclosure of simplifying output to provide "core content" rather than "detailed content" for users farther from the device (Ex. 1003 (Shin) ¶[0089]), it would have been obvious to apply a word replacement approach, such as Iwase's.

149. Sixth, the similarity of the references' outputs would have provided motivation to a person of ordinary skill in the art. Shin discloses an output sentence:

“Prepare your umbrella when you go out.” (Ex. 1003 (Shin) ¶[0088].) Iwase discloses replacing “I recommend” in the output “I recommend bringing an umbrella” with “You should,” yielding: “You should take an umbrella.” (Ex. 1017 (Iwase), 15:6-14.) A person of ordinary skill would have been motivated to modify Shin’s output, simplifying it for users at greater distances, by replacing words in “Prepare your umbrella when you go out” to yield: “Take an umbrella.”

150. Seventh, the similarity of the devices would have motivated a person of ordinary skill in the art to implement Iwase’s word replacement feature, as Iwase (like Shin) determines the user-to-device distance and alters the output based on that distance. (Ex. 1017 (Iwase), 9:59-10:9, 11:32-39, 12:4-9.)

151. A person of ordinary skill in the art would have reasonably expected success in incorporating Iwase’s word replacement feature into Shin’s processor. This configuration would have been trivial, involving straightforward changes to enhance the processor’s approach to simplifying and abbreviating its output content. Furthermore, such an approach was well known in the art, further leading a person of ordinary skill to reasonably expect success. For example, as I discussed above, Ha discloses a device that replaces “team meeting” and “date” (described as “detailed information”) with “schedules” (described as “abstract information”) based on the user’s environment. (Ex. 1011 (Ha) ¶¶[0097]-[0098], Fig. 7; *see generally id.*)

Moreover, Iwase already contemplates incorporation of its word-replacement system into a device like Shin's, which analyzes the user-to-device distance and generates voice output.

152. Thus, Shin and Iwase render this limitation obvious. Because, as I discussed above, Shin discloses the remaining limitations of claim 1, Shin and Iwase render obvious claim 1. (*See* ¶¶49-108, above.)

2. Claim 2

a. 2[a]: Notification Information

153. Claim element 2[a] recites that “the notification information includes type information indicating a type of the content information.” Iwase discloses this element.

154. Iwase's device analyzes the content of notification information to identify clauses containing specific semantic concepts, such as a “request,” “proposal,” “wish,” or “opinion.” (Ex. 1017 (Iwase), 14:35-44; *see also id.*, 14:24-49.) For example, it may detect a clause containing the phrase “I recommend.” (*Id.*, 14:44-45.) The device associates the semantic concept with the phrase in the notification information. (*Id.*, 14:35-40.) The semantic concept is type information because it classifies the meaning of the associated phrase. Thus, Iwase discloses this limitation.

b. 2[b]: Output-Content Generating Unit

155. Claim element 2[b] recites that “the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be 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.” I have been instructed to assume that “placed” is intended to mean “replaced.”

156. Iwase discloses this element. Iwase’s device determines the notification’s subject by identifying a phrase containing a semantic concept. (See ¶154, above.) This semantic concept is type information. Once the subject is determined, words within the subject may be replaced based on user attributes. (Ex. 1017 (Iwase), 15:6-14; see ¶¶143-152, above.) As a result, the device selects a replacement word based on both a word in the notification and its associated type (e.g., semantic concept used to identify subject).

157. A person of ordinary skill in the art would have been motivated to configure Shin’s processor to adjust the information provided to the user based on the user-to-device distance by applying Iwase’s word replacement feature specifically to the information’s subject, and would have reasonably expected success in doing so, for the reasons I discussed above for claim 1. (See ¶¶145-152, above.)

158. Accordingly, for at least the reasons I discuss above, Shin and Iwase render obvious claim 2.

3. Claims 5-7

159. I have discussed above how Shin and Iwase disclose or render obvious the limitations of claims 5-7. (*See* ¶¶49-158, above.) Thus, for at least those reasons, Shin and Iwase render obvious claims 5-7.

C. Claims 1-7 Would Have Been Obvious in View of Shimomura and Aoyama.

160. As I have discussed above, Aoyama discloses certain limitations of claims 1-7 and, with Shin, renders those claims obvious. (*See* ¶¶48-141, above.) The disclosure of Aoyama I discussed above also renders claims 1-2 and 5-7 obvious in combination with Shimomura. To avoid repetition, I will address below only the claims and limitations for which Shimomura is relevant.

1. Claim 1

a. 1[pre]: Output-Content Control Device

161. The preamble of claim 1 recites an “output-content control device.” Shimomura discloses a “speech control device that controls the speech of a device[.]” (Ex. 1004 (Shimomura), claim 1; *see also id.*, Title.) Shimomura discloses an electronic device, such as a robot or other interactive device, with input and output units corresponding to human sensory systems. (*Id.* ¶¶[0020]-[0022], [0096].) Inputs include “a pair of CCD (Charge Coupled Device) cameras 20” (eyes) and “a pair of microphones 21” (ears). (*Id.* ¶[0021].) Outputs include “a speaker 23” (mouth). (*Id.* ¶[0022].) Figure 1 shows an example:

[FIG. 1]

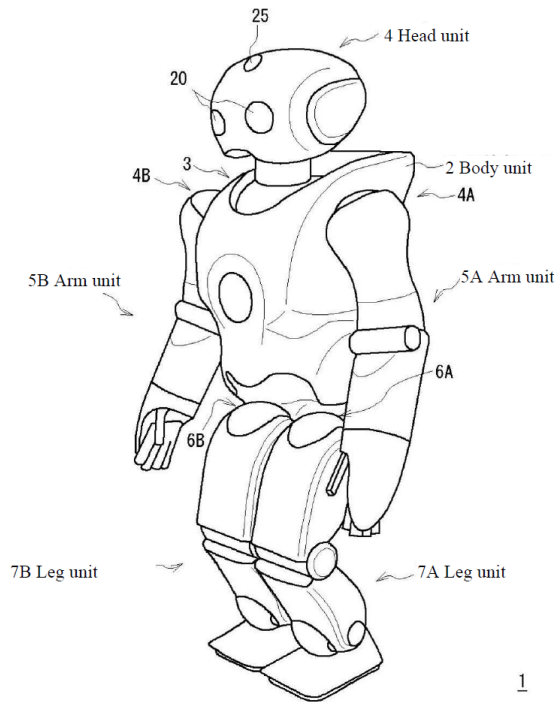


FIG. 1 Hardware constitution of the robot according to this embodiment (1)

(*Id.*, Fig. 1.)

162. The device receives speech via microphones, performs speech recognition, and determines the “content to be spoken[.]” (Ex. 1004 (Shimomura) ¶[0046]; *see also id.*, Abstract, Title, ¶¶[0001], [0030], Fig. 4.) For example, if a user asks about the weather, the device may respond, “Today’s weather in Tokyo is sunny.” (*Id.* ¶[0065].) The device “controls speech forms ... according to the physical distance to the conversation partner.” (*Id.* ¶¶[0049]-[0051]; *see also id.* ¶¶[0009]-[0010], [0016], [0057], [0085], [0088], [0090], [0092], claims 1-24.)

163. Thus, Shimomura discloses an “output-content control device.”

b. 1[a]: Voice Acquiring Unit

164. Claim element 1[a] recites “a voice acquiring unit configured to acquire a voice spoken by a user.” Shimomura discloses this claim element.

165. Shimomura discloses a voice acquiring unit that is configured to “acquire a voice spoken by a user” as recited. Shimomura discloses that its device, robot 1 (orange), includes a control unit 10 (blue) and a microphone 21 (red):

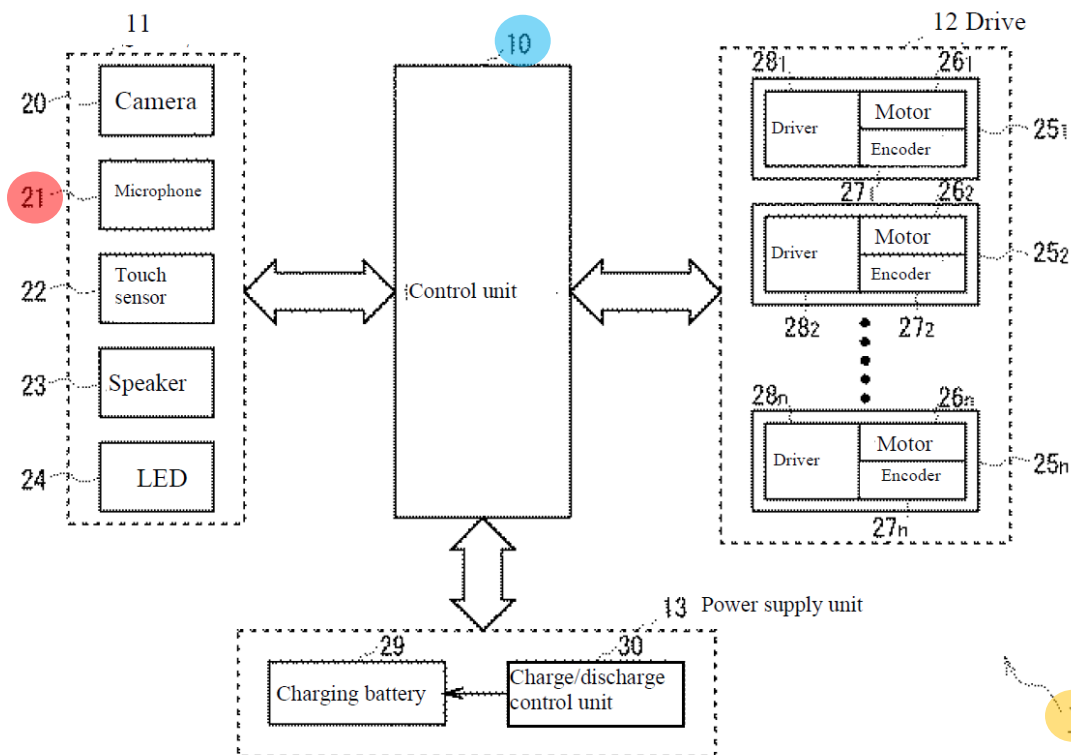


FIG. 2 Hardware constitution of the robot according to this embodiment (2)

(Ex. 1004 (Shimomura), Fig. 2.) The control unit 10, which includes a CPU, controls overall operation and other data processing. (*Id.* ¶¶[0020], [0027], Fig. 3; *see also*

id. ¶¶[0028], [0032], [0034].) The control unit implements a “behavior control system” that includes a “speech recognition unit.” (*Id.* ¶¶[0036]-[0038], [0042], Fig. 4.) “The speech recognition unit 42 performs various sound-related recognition processes such as speech recognition, speaker recognition, and sound source direction recognition based on the speech signal S2 provided by the microphone 21.” (*Id.* ¶[0040].) Thus, Shimomura discloses a voice acquiring unit (speech recognition unit and/or portion of processor that receives speech from microphone) that is configured to acquire a voice spoken by a user.

c. 1[b]: Voice Classifying Unit

166. Claim element 1[b] recites “a voice classifying unit configured to [i] calculate a distance between the user and the output-content control device by a proximity sensor to [ii] classify the voice into either a first voice or a second voice based on the calculated distance.” Shimomura discloses this claim element.

i. 1[b][i]: Calculates Distance

167. Shimomura discloses that its device includes a “distance detection means for detecting the distance to the conversation partner user[.]” (Ex. 1004 (Shimomura) ¶[0093].) Shimomura’s control unit includes an “image recognition unit.” (*Id.* ¶¶[0036]-[0038], [0042], Fig. 4.) The “image recognition unit” may detect and calculate a distance to the user “based on the image signal S1 from the CCD camera [sic] 20.” (*Id.* ¶[0039]; *see also id.* ¶¶[0039] (“distance sensor”), [0059]

(“distance information indicating the distance to the conversation partner detected by the image recognition unit 41” is sent to speech character string transformation unit), [0070].)

168. Thus, Shimomura discloses a voice classifying unit (control unit) configured to calculate a distance between the user and the output-content control device by a proximity sensor (e.g., CCD cameras or other distance sensor).

ii. 1[b][ii]: Classifies Voice Based on Distance

169. Shimomura discloses that its voice classifying unit (processor) analyzes the voice to classify it as either a first or second voice based on the user-to-device distance. Shimomura’s control unit includes a “speech synthesis unit” and a “speech character string transformation unit.” (Ex. 1004 (Shimomura) ¶¶[0036]-[0037], [0046], [0058], [0067], Fig. 4.) These units may “transform the character string” of the content spoken by the device “according to the distance to the conversation partner.” (*Id.* ¶¶[0058], [0067]-[0070]; *see also id.* ¶¶[0060]-[0063], [0073]-[0084], Abstract, claims 1-24.) To do so, Shimomura’s control unit first classifies the voice.

170. For example, Shimomura discloses classifying speech into two categories based on the user-to-device distance: (1) a first voice at greater than 350 cm, and (2) a second voice at less than 350 cm. These classifications are shown in Figure 7:

[FIG. 7]

Second Voice	First Voice
Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

(Ex. 1004 (Shimomura), Fig. 7, ¶¶[0060]-[0064].)

171. Thus, Shimomura discloses the voice classifying unit (processor) is configured to classify the voice into either a first voice (e.g., farther than 350 cm) or a second voice (e.g., closer than 350 cm) based on the calculated distance.

d. 1[c]-1[d]: Intention Analyzing Unit and Notification-Information Acquiring Unit

172. Claim element 1[c] recites “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.” Claim element 1[d] recites “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.” Shimomura and Aoyama disclose or render obvious these claim elements.

173. Shimomura’s control unit is configured to analyze the voice acquired by the voice acquiring unit. (Ex. 1004 (Shimomura) ¶[0040].) It recognizes speech from the microphone’s signal and, based on that speech, determines the content to be spoken to the user. (See ¶¶161-171, above.) Shimomura describes the user as a conversation partner of the device (Ex. 1004 (Shimomura) ¶¶[0010], [0050], [0054], [0057]-[0059], [0067], [0086]-[0087], [0090], claims 1, 9, 17; *see also id.* ¶¶[0003]-[0004]), confirming the device extracts user intent during conversation and responds by performing the required or intended processing.

174. Indeed, Shimomura discloses analyzing speech “to detect intention information” and acquiring “notification information ... based on the intention information” just like the ’675 patent. Shimomura describes analyzing a user’s request to determine intention information (e.g., weather) and then acquiring notification information (e.g., sunny) based on that intention information. (Ex. 1004 (Shimomura) ¶[0065].) Shimomura’s control unit acquires from memory the required notification information, such as “Today’s weather in Tokyo is sunny,” for presenting to the user. (*Id.* ¶¶[0044], [0065].)

175. Even if Shimomura did not expressly disclose these elements, Shimomura’s control unit inherently includes them because they would have been necessary for Shimomura’s device to function as described. These functionalities

are essential for interpreting user input, determining intent, and acquiring appropriate responses. For example, providing weather updates requires interpreting user intent and retrieving relevant content. A person of ordinary skill in the art would have understood that extracting semantic and contextual speech information corresponds to the “intention analyzing unit,” while retrieving relevant content discloses the “notification-information acquiring unit.” These units are required for Shimomura’s control unit to perform as disclosed. Thus, Shimomura expressly or inherently discloses these elements.

176. Even if Shimomura had not disclosed these elements, they would have been obvious to a person of ordinary skill from Shimomura. A person of ordinary skill would have been motivated to use an “intention analyzing unit” to determine user intent and a “notification-information acquiring unit” to retrieve content based on that intent for several reasons.

177. Shimomura describes voice processing capabilities that allow the device to interpret user intent and generate appropriate responses. A person of ordinary skill would have been motivated to structure Shimomura’s control unit with an intention analyzing unit and a notification-information acquisition unit to perform these functions more efficiently. This modular design would facilitate consistent intent extraction, reliable content retrieval, and improved clarity and organization of

internal processes. It would also merely reflect the way known speech analysis systems worked at the time of the alleged invention. (*See* ¶¶33, above; ¶¶232-239, below.)

178. This approach also would have been straightforward, as Shimomura’s device already integrates speech analysis and content generation. Implementing these functionalities as units within Shimomura’s control unit would have been an obvious design choice, enhancing modularity, maintainability, and troubleshooting. This design would predictably improve response accuracy, while offering greater flexibility to handle various user requests. Thus, these claim elements would have been obvious over Shimomura.

179. Lastly, and alternatively, these claim elements were disclosed by Aoyama and the claim as a whole would have been obvious in view of Shimomura and Aoyama. (*See* ¶¶80-84, above.) A person of ordinary skill in the art would have been motivated to implement the intention analyzing unit and notification information acquiring unit from Aoyama for all of the reasons that a person of ordinary skill would have been motivated to similarly modify Shin. (*Id.*)

e. 1[e]: Output-Content Generating Unit

180. Claim element 1[e] recites “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.” Shimomura discloses this limitation.

181. Shimomura discloses an output-content generating unit that generates an output sentence as sentence data to be output to the user. As I discussed above, Shimomura's control unit generates content for output. (See ¶¶166-179, above.) The content can be a sentence including data like: "Today's weather in Tokyo is sunny." (Ex. 1004 (Shimomura) ¶[0065]; *see also id.* ¶¶[0076], [0078]-[0079], [0092].) The control unit includes an "output management unit" and a "behavior selection control unit." (*Id.* ¶¶[0036]-[0037], [0046]-[0047], Fig. 4.) The output management unit "sends the character string provided by the behavior selection control unit 45 to the speech synthesis unit 48, while the speech synthesis unit 48 generates a speech signal S4 of synthetic speech based on the supplied character string and sends it to the speaker 23" for audible presentation. (*Id.* ¶[0047]; *see also id.* ¶¶[0046], [0049]-[0050], [0063], [0083], claims 6, 14, Fig. 13.) Thus, Shimomura discloses an output-content generating unit (control unit) configured to generate an output sentence that is sentence data to be output to the user.

182. Shimomura also discloses the voice-content generating unit generates the output sentence based on the notification information. As I discussed above, Shimomura's control unit determines the content spoken to the user based on the recognized speech so that the device converses with the user. (See ¶¶172-179, above.) Thus, Shimomura discloses this element.

f. 1[f]: Generate a First Output Sentence

183. Claim element 1[f] recites that “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.” Shimomura and Aoyama disclose or render obvious this limitation.

184. Shimomura’s control unit implements the speech character string transformation unit to insert a word like “ne” after each phrase to “chang[e] the interval between phrases so that the division of meaning can be clearly understood according to the physical distance to the conversation partner.” (Ex. 1004 (Shimomura) ¶¶[0057]-[0058]; *see also id.* ¶¶[0059]-[0066], [0085], [0092] (“words other than ‘ne’ can be used”), Figs. 7-8.) When the voice is classified as the first voice (e.g., farther than 350 cm), the device generates a first output by transforming the original content to insert “ne” or another word after phrases. (*Id.* ¶¶[0062], [0065], [0092].) Figure 7 shows when “implementation” of the transformation (e.g., adding “ne”) is performed:

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

First Voice

FIG. 7 Character string transformation determination table

(*Id.*, Fig. 7.) For example, Shimomura discloses “generat[ing] a character string such as ‘Today’s, Tokyo’s, the weather is, sunny’ by adding the character ‘ne’ to the end of each clause-final auxiliary verb, ‘Today’s,’ ‘Tokyo’s,’ and ‘The weather is.’” (*Id.* ¶[0065].)

185. Thus, Shimomura discloses adjusting the output content based on the user-to-device distance, for example to increase user comprehension. As I discussed above, Aoyama discloses replacing words in output content based on the user or situation, for example to adjust how names are presented. (See ¶¶90-108, above.)

186. A person of ordinary skill in the art would have been motivated to combine the disclosures of Shimomura and Aoyama to, for example, adjust Shimomura’s output using Aoyama’s word replacement feature for several reasons.

187. First, Shimomura suggests doing so because it discloses that clear enunciation of content, particularly for distant users, is advantageous. (*E.g.*, Ex. 1004 (Shimomura) ¶¶[0004]-[0017].) A person of ordinary skill would have understood

that one application of Aoyama's word replacement feature would be to replace harder-to-understand formal names with easier-to-understand informal names, motivating a person of ordinary skill to use the feature to advance Shimomura's goals.

188. Second, such a configuration represents the simple addition of one known element (Aoyama's word replacement) to another known element (Shimomura's distance-based output adjustments) to obtain predictable results (output content adjusted based on distance using word replacement). Third, such a configuration reflects applying a known technique (Aoyama's word replacement) to improve a similar device and method (Shimomura's) in the same way (providing easier-to-understand output). Fourth, such a configuration also applies a known technique (Aoyama's word replacement) to a known device and method (Shimomura's content adjustment system) that is ready for improvement and yields predictable results (a system that delivers content to users at greater distances that is adjusted using word replacement).

189. Fifth, it also would have been obvious to combine Shimomura and Aoyama to address the same privacy concerns I discussed above. (*See* ¶99, above.) These concerns would have motivated a person of ordinary skill to incorporate Aoyama's word replacement feature into Shimomura's device to replace words containing personal information with neutral alternatives.

190. A person of ordinary skill in the art would have reasonably expected success in incorporating Aoyama’s word replacement feature into Shimomura’s control unit. This configuration would have been trivial, involving straightforward changes to enhance the control unit’s approach to adjusting output. Furthermore, such an approach was well known in the art, further leading a person of ordinary skill to reasonably expect success. (*See* ¶102, above.)

191. Thus, Shimomura and Aoyama render this limitation obvious.

g. 1[g]: Generate a Second Output Sentence

192. Claim element 1[g] recites that the output-content generating unit is further configured 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.” Shimomura discloses this limitation.

193. Shimomura discloses that the speech character string transformation unit “judges that the character string should not be transformed if the distance to the conversation partner ... is less than 350 [cm].” (Ex. 1004 (Shimomura) ¶¶[0062], [0092]; *see also id.* ¶¶[0057]-[0058], [0065], [0085].) Thus, when the voice is classified as a second voice (e.g., closer than 350 cm), the device generates a second output sentence comprising the original content without adjustment. (*Id.* ¶¶[0062], [0092].) This is shown in Shimomura’s Figure 7:

[FIG. 7] **Second Voice** First Voice

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

(*Id.*, Fig. 7.)

194. For example, Shimomura discloses that the output sentence when the voice is classified as the second voice (e.g., closer than 350 cm) may be: “Today’s weather in Tokyo is sunny.” (Ex. 1004 (Shimomura) ¶[0065].) In this example, the word “weather” conveys all intention information derived from a user statement like “Let me know today’s weather.” The words “Tokyo” and “sunny” represent all content information determined for that user statement. Thus, Shimomura discloses this element.

195. Accordingly, for at least the reasons I discuss above, Shimomura and Aoyama render obvious claim 1.

2. Claims 2-4

196. Aoyama discloses the additional limitations of claims 2-4. (*See* ¶¶109-130, above.) Accordingly, for at least the reasons I discuss above, Shimomura and Aoyama render obvious claims 2-4.

3. Claim 5

197. Claim 5 depends from claim 1 and further recites that “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.” I note that the phrase “the predetermined voice” lacks an antecedent basis; I have been asked to assume that “the predetermined voice” refers to either “the first voice” or “the second voice” as recited in claim 1. Shimomura discloses or renders obvious this limitation.

198. Shimomura discloses that its device receives and processes speech input from a user. (*See* ¶¶164-165, above.) A whisper is a form of speech. Therefore, Shimomura discloses classifying a whispered voice.

199. Even if Shimomura did not teach receiving and processing whispered speech, a person of ordinary skill in the art would have been motivated for several reasons to configure Shimomura’s device to do so.

200. First, the prior art suggests it. For example, Iyer teaches that the privacy concerns I discussed above (*see* ¶99, above) may be addressed in part by processing whispered speech. (Ex. 1012 (Iyer), 3:61-64 (“[W]hen in the discreet mode of operation, the user may whisper voice commands to the microphone.”).) This would have motivated a person of ordinary skill to have Shimomura’s device process whis-

pered speech to address privacy concerns. Furthermore, Shimomura's general disclosure of processing speech (*see* ¶¶161-195, above) would have motivated a person of ordinary skill to process as wide a range of speech as possible, including whispers.

201. Second, such a configuration would represent merely the addition of a known element (processing whispered speech) to another known element (Shimomura's speech processing device) to obtain predictable results (a device capable of adjusting output content in response to whispered speech). Third, such a configuration reflects a combination of known elements (detecting whispers) to achieve an expected result (classifying voice input across wide range of speech, including whispered speech). Fourth, such a configuration reflects applying a known technique (whisper processing) to achieve predictable results (content adjustment based on whispered input).

202. A person of ordinary skill in the art would have reasonably expected success in modifying Shimomura in this way. Extending Shimomura's capabilities to handle whispers would involve straightforward adjustments, if any. Furthermore, such functionality was known in the art, further leading a person of ordinary skill to reasonably expect success. (*See generally, e.g.*, Exs. 1012 (Iyer), 1016 (Schuster); *see* ¶137, above; *see* ¶¶225-231, below.)

203. Accordingly, Shimomura and Aoyama render obvious claim 5.

4. Claims 6 and 7

204. Shimomura discloses both a speech control device and method. (Ex. 1004 (Shimomura) ¶¶[0012]-[0013]; *see also id.* ¶¶[0001], [0009], [0016], [0057], Title, Abstract, Figs. 6, 8, 13, claims 9-16.) Shimomura also discloses electronic devices implemented by instructions stored in a computer-readable storage media. (*Id.* ¶¶[0027], [0030], [0036]; *see also id.* ¶¶[0028]-[0029], [0037]-[0047], [0051]-[0083], Figs. 3-4.) Thus, Shimomura discloses the preamble of claims 6 and 7.

205. As I discuss above, the steps of claims 6 and 7 are identical and merely recite the functions performed by the various units recited in claim 1. (*See* ¶¶139-141, above.) Accordingly, for the reasons I discuss for claim 1, Shimomura and Aoyama render obvious claims 6 and 7. (*See* ¶¶161-195, above.)

D. Claims 1-2 and 5-7 Would Have Been Obvious in View of Shimomura and Iwase.

206. As I have discussed above, Shimomura discloses certain limitations of claims 1-7 and, with Aoyama, renders those claims obvious. (*See* ¶¶160-205, above.) The disclosure of Shimomura I discussed above also renders claims 1-2 and 5-7 obvious in view of Iwase. To avoid repetition, I will address below only the claims and limitations for which Iwase is relevant.

1. Claim 1

207. As I discussed above, Shimomura and Iwase together disclose or render obvious each limitation of claim 1. (*See* ¶¶143-152, 161-195, above.)

208. A person of ordinary skill in the art would have been motivated to configure Shimomura's control unit to adjust the output information provided to the user based on the user-to-device distance by incorporating Iwase's word replacement feature, for example to replace harder-to-understand words with simpler ones when the user is far away, for several reasons.

209. First, Iwase suggests doing so. Iwase discloses that its word replacement system allows a user "to grasp notification contents more effectively." (Ex. 1017 (Iwase), 2:11-14, 15:24-28, 21:66-22:8.) Thus, a person of ordinary skill in the art would have been motivated to incorporate Iwase's word replacement feature into Shimomura's device to make the output content easier to understand.

210. Second, such a configuration represents merely the addition of one known element (Iwase's word replacement) to another known element (Shimomura's distance-based output adjustments) to obtain predictable results (output content adjusted using word replacement). Third, such a configuration reflects applying a known technique (Iwase's word replacement) to improve a similar device and method (Shimomura's content adjustment system) in the same way. Fourth, such a configuration also applies a known technique (Iwase's word replacement) to a known device and method (Shimomura's content adjustment system) that is ready for improvement and yields predictable results (delivering content to users at greater distances that is adjusted using word replacement).

211. A person of ordinary skill would have reasonably expected success in incorporating Iwase’s word replacement feature into Shimomura’s control unit. This configuration would have been trivial, involving straightforward changes to enhance the control unit’s approach to adjusting its output content. Furthermore, such an approach was well known in the art, further leading a person of ordinary skill to reasonably expect success. For example, as I discussed above, Ha discloses a device that replaces “team meeting” and “date” (described as “detailed information”) with “schedules” (described as “abstract information”) based on the user’s environment. (Ex. 1011 (Ha) ¶¶[0097]-[0098], Fig. 7; *see generally id.*)

212. Accordingly, for at least the reasons I discuss above, Shimomura and Iwase render obvious claim 1.

2. Claim 2

213. As I discuss above, Iwase discloses the additional limitations of claim 2. (*See* ¶¶153-158, above.) A person of ordinary skill would have been motivated to configure Shimomura’s control unit to adjust the information provided to the user based on the user-to-device distance by applying Iwase’s word replacement feature at least as to the information’s subject for the reasons I discussed above for claim 1. (*See* ¶¶208-212, above.)

214. Accordingly, for at least the reasons I discuss above, Shimomura and Iwase render obvious claim 2.

3. Claims 5-7

215. I have discussed above how Shimomura and Iwase disclose or render obvious the limitations of claims 5-7. (See ¶¶197-205, above.) Thus, for at least those reasons, Shimomura and Iwase render obvious claims 5-7.

E. Claims 3-4 Would Have Been Obvious in View of (a) Shin, Iwase, and Aoyama; or (b) Shimomura, Iwase, and Aoyama.

216. As I have discussed above, Shin and Iwase render obvious claims 1-2 and 5-7. (See ¶¶142-159, above.) And, as I have also discussed above, Shimomura and Iwase likewise render obvious claims 1-2 and 5-7. (See ¶¶206-215, above.) Furthermore, as I have also discussed above, the additional limitations of claims 3 and 4 are disclosed by Aoyama. (See ¶¶118-130, above.) Thus, claims 3 and 4 would have been obvious in view of (a) Shin, Iwase, and Aoyama; and (b) Shimomura, Iwase, and Aoyama.

1. Claim 3

217. Claim 3 depends from claim 2 and further recites that “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.” Aoyama discloses or renders obvious the additional limitations of claim 3.

218. As I discussed above, Aoyama's device references replacement words from tables, which contain relationship information (e.g., associations between words). (See ¶¶114-122, above.) These tables are stored before generating output sentences, meaning the words they contain are predetermined. (*Id.*) Aoyama's device replaces a word when a word in the notification information matches a word in one of the tables. (See ¶¶93-103, above.)

219. It would have been obvious to configure Shin's processor and/or Shimomura's control unit to implement Iwase's word replacement using tables, as in Aoyama. For example, a person of ordinary skill in the art would have been motivated to configure Shin's processor and/or Shimomura's control unit to use separate tables for storing expression variants corresponding to common semantic concepts disclosed in Aoyama, such as "request," "proposal," "wish," or "opinion," for several reasons.

220. First, such a configuration reflects a combination of known elements (Shin or Shimomura's distance-based output adjustments, Iwase's word replacement, and Aoyama's organization of replacement words in tables) to achieve predictable results (generating responses for distant users by looking up specific words in semantic concept tables and replacing those words with more general ones from the tables). Second, such a configuration reflects applying a known technique (Aoyama's replacement tables) to improve a similar device and method (Shin or

Shimomura's, as modified by Iwase) in the same way. Third, such a configuration applies a known technique (Aoyama's replacement tables) to a known device and method (Shin or Shimomura's, as modified by Iwase) that is ready for improvement and yields predictable results (using tables to replace words for users beyond a distance threshold).

221. A person of ordinary skill would have reasonably expected success in configuring Shin's processor or Shimomura's control unit to implement Iwase's word replacement using tables as in Aoyama. This configuration would have been trivial, involving straightforward changes to enhance the processor's approach to simplifying and abbreviating its output content.

222. Accordingly, for at least the reasons I discuss above, claim 3 would have been obvious over (a) Shin, Iwase, and Aoyama and (b) Shimomura, Iwase, and Aoyama.

2. Claim 4

223. As I discuss above, a person of ordinary skill would have been motivated to combine either Shin or Shimomura with Iwase and Aoyama to use separate tables for storing expression variants corresponding to common semantic concepts (e.g., "request," "proposal," "wish," or "opinion"). (See ¶¶217-222, above.) In this configuration, because each table of expression variants is associated with a distinct semantic concept like "request" or "proposal," each table includes type information

(e.g., “request” or “proposal”). Furthermore, consistent with the disclosure in Iwase and Aoyama, when a word in the notification information has a semantic concept that matches the semantic concept of a table and also appears in that table, the processor would replace the word in the notification with an expression variant from that table. (See ¶¶90-130, 143-158, above.)

224. Accordingly, for at least the reasons I discuss above, claim 4 would have been obvious over (a) Shin, Iwase, and Aoyama and (b) Shimomura, Iwase, and Aoyama.

F. Claim 5 Would Have Been Obvious in View of the Previously Discussed Combinations and Schuster.

225. As I discuss above, claim 5 (and claim 1, from which it depends) would have been obvious over Shin or Shimomura and Aoyama or Iwase. (See ¶¶48-215, above.) Even if those references did not disclose the additional limitation of claim 5, it is disclosed or rendered obvious by Schuster.

226. Claim 5 depends from claim 1 and further recites that “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.” I note that the phrase “the predetermined voice” lacks an antecedent basis; I have been asked to assume that “the predetermined voice” refers to either “the first voice” or “the second voice” as recited in claim 1.

227. Schuster discloses an electronic device (e.g., mobile communication device) with a processor that interacts with a user by receiving verbal queries and providing voice outputs like search answers or recommendations. (Ex. 1016 (Schuster), Abstract, 1:7-11, 3:31-39, 7:18-20.) Schuster’s device “enters a private communication mode with the user when the user’s preceding voice interaction with the device is whispered.” (*Id.*, 2:59-62.) The private mode alters the device’s communication behavior by, for example, “significantly lower[ing] the volume of the speech” (*id.*, 3:31-39), which ensures privacy, prevents social awkwardness, and maintains confidentiality. (*Id.*, Abstract, 2:62-67.) Because Schuster’s device detects whispered input through microphone signal analysis (*Id.*, 3:2-4) and, upon detecting a whisper, flags the interaction as private or shifts into private mode (*id.*, 3:12-18; *see also id.*, 2:6-10, 6:34-46, claims 3-4, 12-13, 19), Schuster discloses distinguishing between a first voice and a second voice based on whether the voice is whispered or non-whispered. Thus, even if claim 5 were to require the voice to be classified based on whether it is a whisper, the claim would have been obvious in view of the references I previously discussed and further in view of Schuster.

228. A person of ordinary skill in the art would have been motivated for several reasons to configure Shin’s processor and/or Shimomura’s control unit to process whispered speech and/or classify a voice based on whether it is a whisper.

229. First, such a configuration represents merely the simple addition and/or substitution of one or more known elements (Schuster's processing of whispers and/or classification of voices based on detected whispers) to and/or for another known element (Shin or Shimomura's processing and classification of voices) to obtain predictable results (a system that classifies speech, including whispers, based on distance and/or detected whispers). Second, such a configuration reflects nothing more than applying a known technique (Schuster's processing of whispers and/or classification of voices based on detected whispers) to improve a similar device and method (Shin's or Shimomura's) in the same way. Third, such a configuration also applies a known technique (Schuster's processing of whispers and/or classification of voices based on detected whispers) to a known device and method (Shin's or Shimomura's) that is ready for improvement and yields predictable results (a system that classifies speech, including whispers, based on distance and/or detected whispers). Fourth, Schuster discloses classifying based on distance (Ex. 1016 (Schuster), 3:27-30), just as Shin and Shimomura do.

230. A person of ordinary skill would have reasonably expected success in modifying Shin and/or Shimomura in this way. Shin and Shimomura's devices already classify speech into two or more categories. Extending their capabilities to process whispers and/or classify users based on whispered and non-whispered input

would involve straightforward adjustments (if any) to the existing voice detection and processing hardware or software.

231. Accordingly, for at least the reasons I discuss above, Shin or Shimomura and Aoyama or Iwase, in further view of Schuster, render obvious claim 5.

G. Claims 1-7 Would Have Been Obvious in View of the Previously Discussed Combinations and Kristjansson.

232. As I discuss above, claims 1-7 would have been obvious over Shin or Shimomura, Aoyama and/or Iwase, and/or Schuster. (See ¶¶48-231, above.) Even if those references did not render the claims obvious, the relevant limitations are disclosed or rendered obvious by Kristjansson. Specifically, even if the other references did not disclose the “intention analyzing unit” and “notification-information acquiring unit” limitations (e.g., claim elements 1[c]-1[d], 6[c]-6[d], and 7[c]-7[d]), Kristjansson discloses them and renders the claims obvious.

233. For example, claim element 1[c] recites “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.” Claim element 1[d] recites “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.” These

limitations refer to determining the user’s intent using natural language understanding (“NLU”) processing and acquiring the desired information. (Ex. 1001 (’675 patent), 4:25-27.)

234. Kristjansson discloses a “speech processing system” that, similar to those disclosed in Shin and Shimomura, comprises “one or more speech-controlled devices 110 local to user 5.” (Ex. 1013 (Kristjansson), 2:45-67.) Kristjansson’s system is “capable of performing traditional speech processing” and “may execute certain commands, such as answering spoken utterances of the users[.]” (*Id.*) Kristjansson discloses that to do so, the system includes “an NLU component” that “takes textual input (such as processed from ASR [automatic speech recognition] 450 based on the utterance input audio 11 [from the user]) and ... derive[s] an intent or a desired action from the user[.]” (*Id.*, 7:26-67.) Upon determining the user’s intent, the system sends that intent information to a “command processor 490,” which retrieves content based on that intent. (*Id.*, 10:38-11:12.) For example, the command processor may retrieve a song requested by the user (*id.*, 10:44-48), results of a user’s search (*id.*, 10:49-56), or content from another source requested by the user (*id.*, 10:56-11:12). In each case, the retrieved content is delivered to the user. (*Id.*, 10:38-11:12.) Thus, Kristjansson discloses elements 1[c]-1[d]. For the same reasons, Kristjansson discloses the equivalent limitations in the other claims, i.e., claim elements 6[c]-6[d] and 7[c]-7[d].

235. As I discussed above, Shin and Shimomura, like Kristjansson, disclose devices that receive voice input from a user and return voice output. (*See* ¶¶49, 161-162, above.) A person of ordinary skill in the art would have been motivated to implement Kristjansson's NLU teachings in Shin and/or Shimomura's devices to process the user's input and determine and retrieve the proper output for several reasons.

236. First, Shin and Shimomura suggest doing so. Shin and Shimomura disclose devices that analyze voice input, identify content the user is interested in, and generate the requested content. Kristjansson provides additional detail about how to structure such a device, with NLU processing and content retrieval. Thus, a person of ordinary skill would have looked to Kristjansson when implementing Shin and/or Shimomura's devices.

237. Second, Kristjansson's disclosure represents a routine and well-established approach to providing a user with requested content. (*See* Ex. 1015 (Duta), 431-33.) Third, such a modification merely applies a known technique (Kristjansson's NLU processing and content retrieval) to achieve predictable results (providing a user with requested content). Fourth, such a modification reflects merely applying a known technique (Kristjansson's NLU processing) to a known device (Shin's and/or Shimomura's) that is ready for improvement and yields predictable results (providing a user with requested content).

238. A person of ordinary skill would have reasonably expected success in integrating Kristjansson's teachings with Shin's and/or Shimomura's at least because Kristjansson's approach was well-known in the art (*e.g.*, Ex. 1015 (Duta), 431-33) and because Shin and Shimomura contemplate modification or addition to their systems by persons of ordinary skill in the art (*e.g.*, Ex. 1003 (Shin) ¶[0184]; Ex. 1004 (Shimomura) ¶[0089]). Moreover, Kristjansson discloses a system, like Shin and Shimomura, that uses a proximity sensor to determine the user-to-device distance and adjusts the output format (*e.g.*, volume) based on that distance. (Ex. 1013 (Kristjansson), claim 1, 3:1-27.)

239. Accordingly, for at least the reasons I discuss above, Shin or Shimomura, Aoyama and/or Iwase, and/or Schuster, in further view of Kristjansson, render obvious claims 1-7.

VII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

240. I am not aware of any secondary considerations of nonobviousness. If Patent Owner identifies any alleged evidence of secondary considerations in the future, I reserve the right to respond to that information.

VIII. CONCLUSION

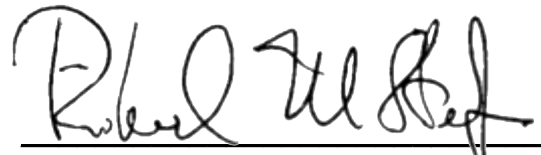
241. For the foregoing reasons, it is my opinion that claims 1-7 of the '675 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention in view of, the prior art discussed above.

242. I reserve the right to supplement my opinions in the future to address or respond to any issues that the Patent Owner may raise, as well as new information including, but not limited to, any claim constructions advanced by the Patent Owner or adopted by the Board in the Institution Decision, and respond to any alleged secondary considerations as they become available to me.

Amazon.com, Inc. v. SoundClear Technologies LLC
Declaration of Richard Stern, Ph.D. – Patent 11,244,675

I declare that all statements made herein of my 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 of Title 18 of the United States Code.

Executed on February 25, 2025 at Pittsburgh, Pennsylvania.



Richard Stern, Ph.D.