

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. IPR2025-00673
Patent No. 11,244,675

**PETITION FOR *INTER PARTES* REVIEW
OF CLAIMS 1-7 OF U.S. PATENT NO. 11,244,675**

TABLE OF CONTENTS

I.	INTRODUCTION -----	1
II.	BACKGROUND -----	2
III.	THE '675 PATENT -----	3
	A. Overview -----	3
	B. Prosecution History -----	5
IV.	RELIEF REQUESTED -----	5
	A. Grounds -----	5
	B. Prior Art -----	6
V.	LEVEL OF ORDINARY SKILL -----	7
VI.	CLAIM CONSTRUCTION -----	7
VII.	GROUND 1A: CLAIMS 1-7 WOULD HAVE BEEN OBVIOUS OVER SHIN AND AOYAMA -----	7
	A. Claim 1 -----	7
	1. 1[pre]: Output-Content Control Device -----	7
	2. 1[a]: Voice Acquiring Unit -----	8
	3. 1[b]: Voice Classifying Unit -----	10
	a. 1[b][i]: Calculates Distance -----	10
	b. 1[b][ii]: Classifies Voice Based on Distance -----	12
	4. 1[c]-1[d]: Intention Analyzing Unit and Notification- Information Acquiring Unit -----	15
	5. 1[e]: Output-Content Generating Unit -----	20
	6. 1[f]: First Output Sentence -----	22
	7. 1[g]: Second Output Sentence -----	28

B.	Claim 2 -----	30
1.	2[a]: Notification Information -----	30
2.	2[b]: Output-Content Generating Unit-----	32
C.	Claim 3 -----	33
D.	Claim 4 -----	35
1.	4[a]: Relationship Information-----	35
2.	4[b]: Output-Content Generating Unit-----	36
E.	Claim 5 -----	37
F.	Claims 6-7-----	40
VIII.	GROUND 1B: CLAIMS 1-2 AND 5-7 WOULD HAVE BEEN OBVIOUS OVER SHIN AND IWASE -----	41
A.	Claim 1 -----	41
B.	Claim 2 -----	44
1.	2[a]: Notification Information -----	44
2.	2[b]: Output-Content Generating Unit-----	45
C.	Claims 5-7-----	46
IX.	GROUND 2A: CLAIMS 1-7 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA AND AOYAMA -----	46
A.	Claim 1 -----	46
1.	1[pre]: Output-Content Control Device -----	46
2.	1[a]: Voice Acquiring Unit-----	48
3.	1[b]: Voice Classifying Unit-----	49
a.	1[b][i]: Calculates Distance -----	49
b.	1[b][ii]: Classifies Voice Based on Distance-----	49

4.	1[c]-1[d]: Intention Analyzing Unit and Notification- Information Acquiring Unit-----	51
5.	1[e]: Output-Content Generating Unit-----	53
6.	1[f]: First Output Sentence -----	54
7.	1[g]: Second Output Sentence -----	57
B.	Claims 2-4-----	58
C.	Claim 5 -----	58
D.	Claims 6-7-----	60
X.	GROUND 2B: CLAIMS 1-2 AND 5-7 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA AND IWASE -----	60
A.	Claim 1 -----	60
B.	Claim 2 -----	62
C.	Claims 5-7-----	62
XI.	GROUND 3: CLAIMS 3-4 WOULD HAVE BEEN OBVIOUS OVER THE COMBINATIONS IN GROUNDS 1B OR 2B AND AOYAMA -----	63
A.	Claim 3 -----	63
B.	Claim 4 -----	64
XII.	GROUND 4: CLAIM 5 WOULD HAVE BEEN OBVIOUS OVER THE COMBINATIONS IN GROUNDS 1A-2B AND SCHUSTER -----	65
XIII.	GROUND 5: CLAIMS 1-7 WOULD HAVE BEEN OBVIOUS OVER THE COMBINATIONS IN GROUNDS 1A-4 AND KRISTJANSSON-----	67
XIV.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	70
XV.	DISCRETIONARY DENIAL UNDER §314(A) IS NOT APPROPRIATE -----	70

A.	Petitioner’s <i>Sotera</i> Stipulation-----	71
B.	Compelling Evidence of Unpatentability -----	71
XVI.	DISCRETIONARY DENIAL UNDER §325(D) IS NOT APPROPRIATE -----	72
XVII.	MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT-----	73
A.	Real Party-In-Interest (37 C.F.R. §42.8(b)(1))-----	73
B.	Related Matters (37 C.F.R. §42.8(b)(2))-----	73
C.	Lead and Backup Counsel (37 C.F.R. §42.8(b)(3)) -----	74
D.	Service Information (37 C.F.R. §42.8(b)(4))-----	74
E.	Grounds for Standing (37 C.F.R. §42.104)-----	75
F.	Payment of Fees (37 C.F.R. §42.15(a)) -----	75

TABLE OF AUTHORITIES

<i>Cases:</i>	<i>Page(s):</i>
<i>Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH,</i> IPR2019-01469, Paper 6 (P.T.A.B. Feb. 13, 2020) -----	72, 73
<i>Apple Inc. v. Telefonaktiebolaget LM Ericsson,</i> IPR2022-00457, Paper 7 (P.T.A.B. Sept. 21, 2022) -----	73
<i>Endymed Med. Ltd. v. Serendia, LLC,</i> IPR2024-00843, Paper 14 (P.T.A.B. Jan. 10, 2025) -----	71
<i>JUUL Labs, Inc. v. NJOY, LLC,</i> IPR2024-00160, Paper 10 (P.T.A.B. May 24, 2024) -----	72
<i>KSR Int’l Co. v. Teleflex Inc.,</i> 550 U.S. 398 (2007) -----	passim
<i>Leapfrog Enters. v. Fisher-Price, Inc.,</i> 485 F.3d 1157 (Fed. Cir. 2007) -----	70
<i>Newell Cos. v. Kenney Mfg. Co.,</i> 864 F.2d 757 (Fed. Cir. 1988) -----	70
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.,</i> 868 F.3d 1013 (Fed. Cir. 2017) -----	7
<i>Quasar Sci. LLC v. Colt Int’l Clothing, Inc.,</i> IPR2023-00611, Paper 10 (P.T.A.B. Oct. 10, 2023) -----	72, 73
<i>Shenzen Chic Elecs. v. Pilot, Inc.,</i> IPR2023-00810, Paper 12 (P.T.A.B. Nov. 8, 2023) -----	72
<i>Sotera Wireless, Inc. v. Masimo Corp.,</i> IPR2020-01019, Paper 12 (P.T.A.B. Dec. 1, 2020) -----	71
<i>TP-Link Corp. Ltd. v. Netgear, Inc.,</i> IPR2023-01469, Paper 10 (P.T.A.B. Apr. 2, 2024) -----	72
<i>Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.,</i> 200 F.3d 795 (Fed. Cir. 1999) -----	7

Statutes and Rules:

35 U.S.C. §102-----	6
35 U.S.C. §103-----	5
35 U.S.C. §112-----	7
35 U.S.C. §314-----	70, 71
35 U.S.C. §325-----	72

Miscellaneous:

Katherine K. Vidal, <i>Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation</i> (June 21, 2022) -----	71
--	----

LIST OF EXHIBITS

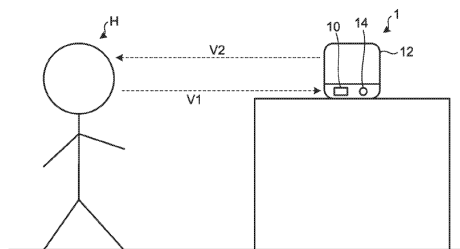
Exhibit	Description
1001	U.S. Patent No. 11,244,675
1002	Declaration of Richard Stern, Ph.D.
1003	U.S. Patent App. Publ. No. 2017/0083281 (“Shin”)
1004	English Translation of Shimomura from Ex. 1005
1005	Declaration of Gwen Snorteland for Translation of Japanese Unexamined Patent App. Publ. 2005/202076 (“Shimomura”)
1006	Curriculum Vitae of Richard Stern, Ph.D.
1007	Excerpts from File History of U.S. Patent No. 11,244,675
1008	Order (Dkt. No. 63), <i>SoundClear Techs., LLC v. Amazon.com, Inc.</i> , No. 1:24-cv-01283-AJT-WBP (E.D. Va. Nov. 8, 2024)
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”)
1014	Order (Dkt. No. 84), <i>SoundClear Techs., LLC v. Amazon.com, Inc.</i> , No. 1:24-cv-01283-AJT-WBP (E.D. Va. Jan. 10, 2025)
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”)

Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively, “Petitioner” or “Amazon”) request *inter partes* review of claims 1-7 of U.S. Patent No. 11,244,675 (“the ’675 patent”), which SoundClear Technologies LLC (“Patent Owner” or “PO”) purportedly owns.

I. INTRODUCTION

The ’675 patent describes an electronic device (e.g., speaker or phone) that receives and processes a user’s voice input V1 and outputs a response voice V2:

FIG.1



(Ex. 1001, Fig. 1, 2:64-3:4.) The patent claims priority to 2018, years after Apple (2011), Amazon (2014), and Google (2016) launched their voice assistants. The patent admits that devices that detect voice, perform processing according to the user’s intent, and provide voice output were known. (*Id.*, 1:20-27.)

The Examiner allowed the claims because they recite calculating the user-to-device distance and tailoring the response based on that distance. (Ex. 1007, 35-36.) But many prior art references disclosed this. Because the ’675 patent claims would have been obvious to those skilled in the art, the Board should cancel the claims.

II. BACKGROUND

Determining a user’s distance from a device and tailoring output based on that distance has been known for decades. (Ex. 1002 ¶¶33-43.) In 2005, Shimomura (filed by Sony) described doing so. (Ex. 1004, Abstract.) Shimomura used cameras to determine the user-to-device distance, classified the user’s speech based on distance (e.g., more or less than 350cm away) and adjusted the output content accordingly (*id.* ¶¶[0039], [0085]).

In 2017, Shin (filed by Samsung) disclosed a device that uses a “distance detection module” or “proximity sensor” to “compute a distance between a user and the electronic device[.]” (Ex. 1003 ¶¶[0066], [0153].) Shin’s device tailored the output based on that distance. For example, for users within 1 meter, the device output “detailed content”; for users more than 1 meter away, it output “abbreviated content.” (*Id.* ¶¶[0051], [0088], Table 3.)

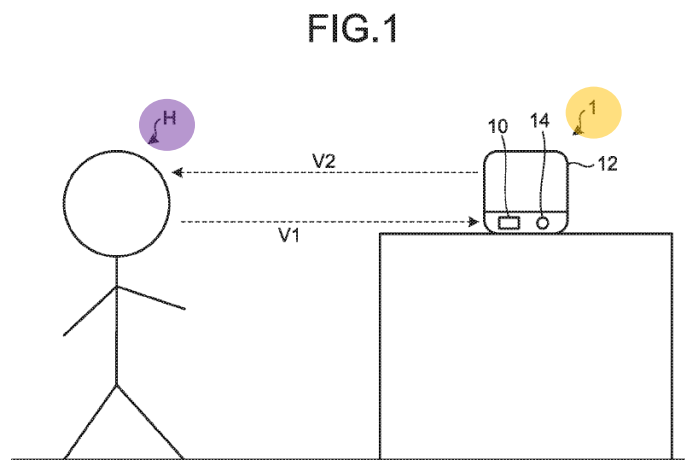
Tailoring output by replacing words was also known, as the Examiner acknowledged. (Ex. 1007, 35-36, 63-70, 98-104; *see also* Ex. 1002 ¶43 (citing Ex. 1018).) In 2017, Aoyama (filed by Sony) disclosed customizing a device’s voice output by replacing words based on the user or situation. (Ex. 1010 ¶¶[0138]-[0140], [0177], [0218].) For example, the device could replace names such as “Taro Yamada” with “Mr. Yamada” in its output. (*Id.* ¶¶[0138]-[0140]; Ex. 1002 ¶41.)

Iwase (filed by Sony) also described a mobile device that tailored voice output by replacing words. (Ex. 1017, 3:26-29, 5:10-13, 15:6-14, Abstract.)

III. THE '675 PATENT

A. Overview

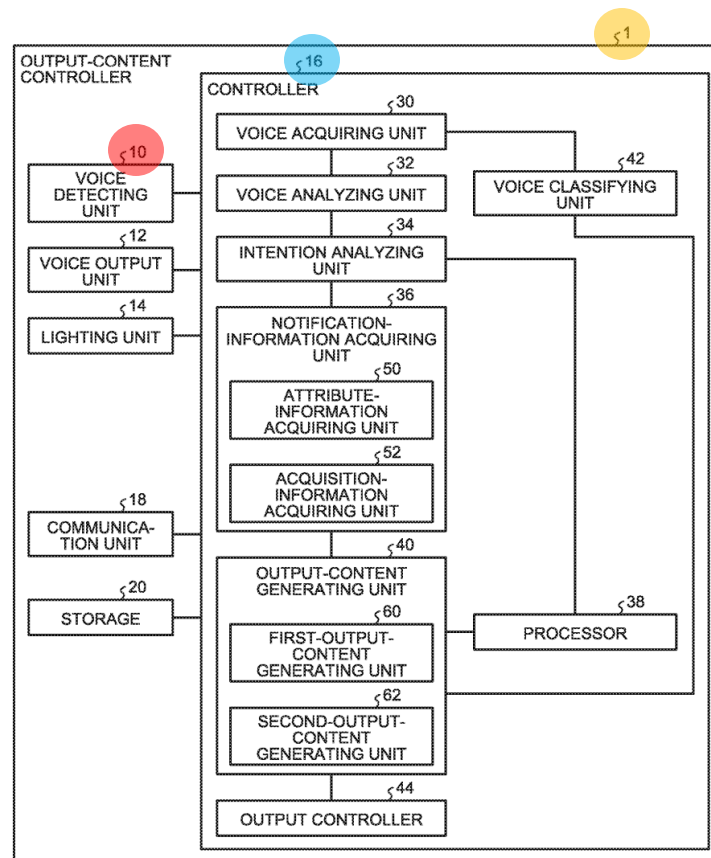
The '675 patent describes a device that analyzes a user's voice and generates a response. (Ex. 1001, Abstract.) Figure 1 shows such a device 1 (orange) that detects the voice V1 of a user H (purple), processes it, and outputs a responsive voice V2:



(*Id.*, Fig. 1, 2:64-3:4; Ex. 1002 ¶44.)¹ The device includes various components, such as a voice detecting unit 10 (red) and a controller 16 (blue):

¹ Figures and Tables herein may be colored or annotated for clarity.

FIG.2



(Ex. 1001, Fig. 2, 3:17-22.) The controller includes “units” for acquiring the voice, analyzing it, processing the request, and generating a response. (*Id.*, 3:55-67.)

The device may include a “proximity sensor” for calculating the distance to the user. (*Id.*, 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) with another word (e.g., Yoshi), and output the 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 using the original word. (*Id.*; *id.*, 11:5-52; Ex. 1002 ¶¶45-46.)

B. Prosecution History

During prosecution, the applicant distinguished the prior art based on limitations that recite calculating the user-to-device distance and classifying the voice based on distance. (Ex. 1007, 47, 49-50.) Apparently unaware that prior art disclosed these limitations, the Examiner erroneously allowed the claims. (*Id.*, 35-36.)

IV. RELIEF REQUESTED

A. Grounds

The Board should cancel the claims under 35 U.S.C. §103 on the following Grounds:

Ground	Reference(s)	Challenged Claims
1a	Shin and Aoyama	1-7
1b	Shin and Iwase	1-2, 5-7
2a	Shimomura and Aoyama	1-7
2b	Shimomura and Iwase	1-2, 5-7
3	Grounds 1b or 2b and Aoyama	3-4
4	Grounds 1a-2b and Schuster	5
5	Grounds 1a-4 and Kristjansson	1-7

This Petition is supported by the expert declaration of Richard Stern (Exs. 1002, 1006).

B. Prior Art

The '675 patent's earliest possible priority date is March 12, 2018. (Ex. 1001.) The references relied on herein are prior art for the following reasons:

Ex.	Reference	Date	Art Type
1003	Shin	Filed: September 19, 2016 Published: March 23, 2017	§102(a)(2) §102(a)(1)
1004	Shimomura	Published: October 28, 2005	§102(a)(1)
1010	Aoyama	Filed: November 26, 2015 Published: November 23, 2017	§102(a)(2) §102(a)(1)
1013	Kristjansson	Filed: March 30, 2017	§102(a)(2)
1016	Schuster	Filed: June 16, 2016 Published: June 13, 2017	§102(a)(2) §102(a)(1)
1017	Iwase	Filed: December 25, 2017	§102(a)(2)

Shimomura published as a Japanese Patent Application. A certified English translation (Ex. 1004) is relied on herein. (*See* Ex. 1005.)

The references above are analogous art because they are from the same field as the '675 patent, e.g., controlling a device's voice output. (Ex. 1001, 1:28-61, 3:10-16). They are also pertinent to a problem the inventors were focused on, e.g., adjusting voice output to improve user interaction. (Ex. 1002 ¶23.)

V. LEVEL OF ORDINARY SKILL

A person of ordinary skill in the art (“POSITA”) 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. (*Id.* ¶¶28-32.) Work experience could substitute for formal education and additional formal education could substitute for work experience. (*Id.*)

VI. CLAIM CONSTRUCTION

No claim terms require construction to resolve the invalidity challenges here. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999). For purposes of this proceeding only, Petitioner assumes the claims are not invalid under §112.

VII. GROUND 1A: CLAIMS 1-7 WOULD HAVE BEEN OBVIOUS OVER SHIN AND AOYAMA.

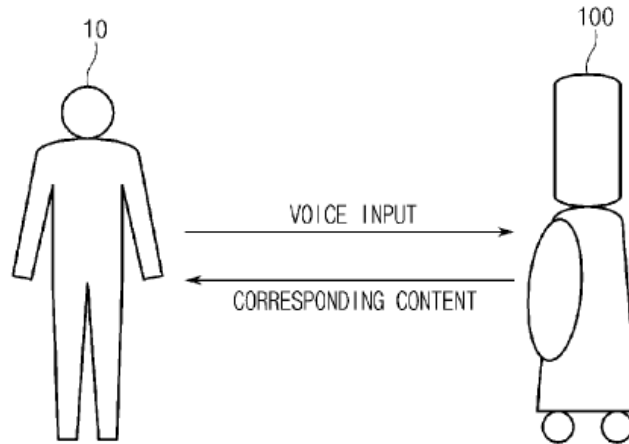
Shin and Aoyama render obvious claims 1-7. (Ex. 1002 ¶¶48-141.)

A. Claim 1

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

Shin discloses a device that receives a user’s voice input, “generate[s] content corresponding to a result of analyzing the voice input,” and “provide[s] the generated content as sound[.]” (Ex. 1003 ¶[0037]; *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].) Figure 1A shows an example of device 100:



(*Id.*, Fig. 1A.) The device controls the output content. (*Infra* §§VII.A.5-VII.A.7.) Thus, Shin discloses an “output-content control device.” (Ex. 1002 ¶49.)

2. 1[a]: Voice Acquiring Unit

Claim element 1[a] recites “a voice acquiring unit configured to acquire a voice spoken by a user.” The ’675 patent states that a voice acquiring unit may be part of a controller, which can be a central processing unit (CPU). (Ex. 1001, 3:55-67, Fig. 2.) Thus, the “voice acquiring unit” refers to a CPU configured to perform the claimed function, namely, to acquire a voice spoken by a user. (Ex. 1002 ¶51.)

Shin discloses this limitation. (*Id.* ¶¶50-53.) Shin’s device 101 (orange) includes a processor 120 (blue) and an audio input module 151 (red):

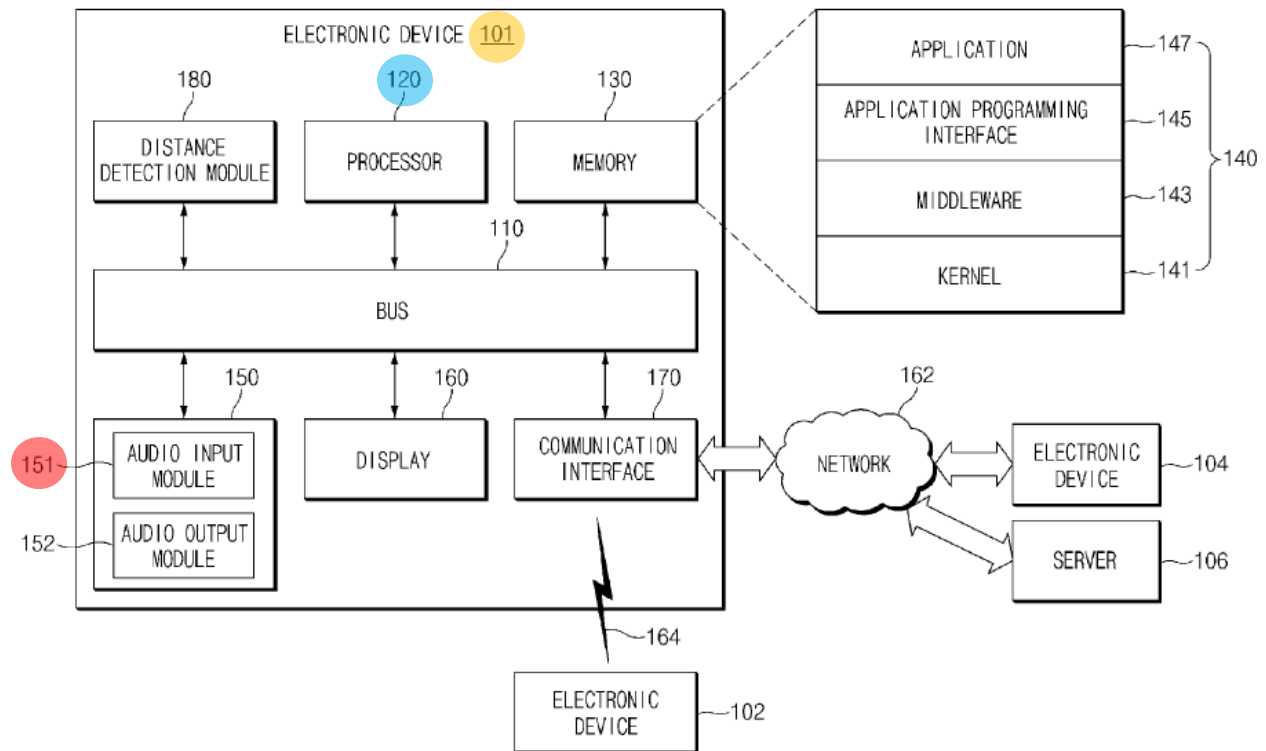


FIG. 3

(Ex. 1003, 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]; *id.*, Abstract, ¶¶[0003]-[0011], [0037]-[0039], [0041], [0060]-[0061], claims 1, 15, 20.) The audio input module 151 can be “implemented with a microphone” to “obtain a user’s speech as a voice input.” (*Id.* ¶¶[0062]; *id.* ¶¶[0037], [0155], [0158].)

Thus, Shin discloses a voice acquiring unit (e.g., audio input module and/or portion of processor that receives speech) configured to acquire a voice spoken by a user. (Ex. 1002 ¶¶50-53.)

3. 1[b]: Voice Classifying Unit

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.” The “voice classifying unit” refers to a CPU configured to perform the claimed function. (*See* Ex. 1001, 3:55-67, Fig. 2; Ex. 1002 ¶¶55.) Shin discloses this claim element. (Ex. 1002 ¶¶54-70.)

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

Shin’s processor “compute[s] the distance between the user and the electronic device[.]” (Ex. 1003 ¶¶[0066].) This may be performed based on image data obtained from a distance detection module. (*Id.*; *see also id.* ¶¶[0053] (“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.) The module may include a proximity sensor, such as “a depth camera” and/or “various sensors,” such as “an infra-red sensor, an RF sensor, an ultrasonic sensor, and the like.” (*Id.* ¶[0066]; *id.* ¶[0042].) Shin’s Figure 3 shows the device 101 comprising processor 120 (blue) and distance detection module 180 (green):

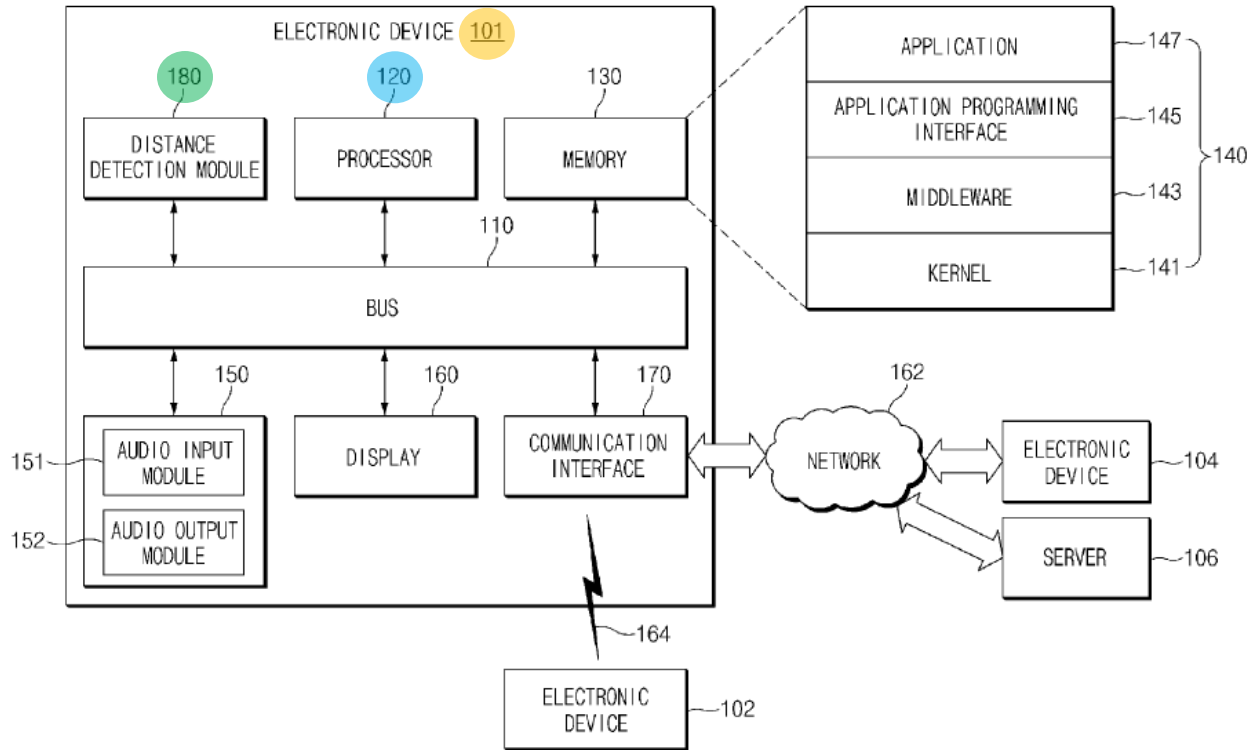


FIG. 3

(*Id.*, Fig. 3, ¶¶[0045]; *id.* ¶¶[0053]-[0054], [0077], [0096], claims 5, 9; Ex. 1002 ¶56.)²

Shin also expressly discloses that the electronic device 101 may use “a proximity sensor.” (Ex. 1003 ¶[0153], Fig. 9 (“proximity sensor” 940G in “sensor module” 940), ¶[0143] (device 901 may be included in devices 100, 101); Ex. 1002 ¶57.)

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. (Ex. 1002 ¶¶56-58.)

² Shin’s electronic device 100 may be implemented with the modules of electronic device 101. (Ex. 1003 ¶¶[0041], [0043].)

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

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, 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.

Shin’s processor analyzes the voice to classify it as either a first or second voice based on the user-to-device distance. (Ex. 1002 ¶60.) Shin’s processor executes a speech recognition application to process the speech input and generates corresponding output content. (Ex. 1003 ¶¶[0047], [0060]; *id.* ¶¶[0037], [0039].) The processor determines the “output scheme,” including the content, “based on the distance between the user and the electronic device 101[.]” (*Id.* ¶¶[0078], [0086]; *id.* ¶¶[0053], [0077], claim 17.) To do so, Shin’s processor classifies the voice based on the distance. (Ex. 1002 ¶60.)

As shown in Table 3, Shin’s device provides 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, Table 3 (consolidated).) Shin thus 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:

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.*; *id.* ¶¶[0051] (content may be “classified dichotomously”), [0086]-[0087]; Ex. 1002 ¶61.) Any distance range beyond 1 meter would satisfy the claimed “first voice.” (Ex. 1002 ¶61.) Alternatively, all distances over 1 meter could collectively be a first voice (e.g., a voice greater than 1 meter away). (*Id.*)

Thus, Shin discloses that the voice classifying unit (processor) is configured to classify the voice as either a first voice (e.g., 1-2 meters) or a second voice (e.g.,

<1 meter) based on the calculated distance. (Ex. 1002 ¶¶59-62.)

Even if PO were to argue that this claim element requires no more than two classifications, Shin would render this obvious. (*Id.* ¶¶63-70.) A POSITA would have been motivated to implement Shin’s amount-of-information table (Table 3) with only two distance ranges for several reasons.

First, Shin suggests it. Shin discloses dividing content into two classifications. (Ex. 1003 ¶[0051].) Consequently, a POSITA would have been motivated to also use two distance ranges, one for each classification. (Ex. 1002 ¶64.)

Second, Shin is “not limited to the example[] of ... Table 3.” (Ex. 1003 ¶[0090].) This flexibility would have motivated a POSITA to implement a simpler, two-classification system—such as “near” and “far”—sufficient for applications requiring only basic distinctions. (Ex. 1002 ¶65.)

Third, simplifying complex systems into fewer categories is a routine and well-established approach to reduce computational load and improve efficiency. (*Id.* ¶66); *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007). Binary classification (having two categories) is a standard approach in many fields, including voice-controlled devices (*see, e.g.*, Ex. 1004, Figs. 7, 12), to simplify user experience. (Ex. 1002 ¶66.) Many consumer devices adopt binary distinctions, such as “low” and “high” or “near” and “far.” (*Id.*) A POSITA would have recognized binary classification as an efficient, user-friendly solution for Shin’s needs. (*Id.*)

Fourth, such a modification reflects merely applying a known technique (binary classification) to achieve predictable results (content output adjusted based on “near” or “far” proximity classifications). (*Id.* ¶67); *KSR*, 550 U.S. at 417.

Fifth, such a modification reflects applying a known technique (binary classification) to a known device (Shin’s) that is ready for improvement and yields predictable results (adjustments to content detail based on proximity threshold). (Ex. 1002 ¶68); *KSR*, 550 U.S. at 417.

A POSITA would have reasonably expected success in modifying Shin in this way because doing so would have been trivial, involving configuring Shin’s processor with revised tables for two proximity-based categories. (*Id.* ¶69.)

4. 1[c]-1[d]: Intention Analyzing Unit and Notification-Information Acquiring Unit

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.”

The intention analyzing unit and notification-intention acquiring unit may be parts of a controller or CPU. (Ex. 1001, 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., 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 performing these functions was known (*id.*, 1:20-27; Ex. 1002 ¶72) and, consequently, these limitations cannot make the claim patentable. Regardless, Shin discloses and renders obvious these claim elements. (Ex. 1002 ¶¶71-79.)

Shin’s processor analyzes the voice acquired by the voice acquiring unit. (Ex. 1003 ¶¶[0048], [0062].) Shin’s processor executes a speech recognition application to process the speech input and generate corresponding content for output by analyzing the request. (*Id.* ¶¶[0048], [0058]-[0060], [0084], [0089], [0037], [0058], [0067], [0071], [0088], [0103], [0107], [0120], [0126]-[0127], [0141]; Ex. 1002 ¶73.)

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. (Ex. 1002 ¶74.) 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. (*Id.*; Ex. 1003 ¶[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. (Ex. 1003 ¶¶0037]; *id.* ¶¶0039]; Ex. 1002 ¶74.)

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 ¶¶0060].) The device can also delegate some processing to another device (e.g., server), which delivers the results. (*Id.* ¶¶0067]; *id.* ¶¶0064]-[0065], [0174]-[0176].) Accordingly, Shin discloses acquiring the information from memory or from another device. (Ex. 1002 ¶75.)

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. (*Id.* ¶76.) 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”). (*Id.*) Thus, Shin expressly or inherently discloses these elements. (*Id.* ¶¶71-76.)

Even if Shin had not disclosed these claim elements, they would have been obvious over Shin. (*Id.* ¶¶77-79.) Shin’s processor analyzes voice input to identify

content of interest, then acquires and provides that content. A POSITA 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. (*Id.* ¶77.) Such a modular design would have facilitated consistent intent extraction and reliable content retrieval, and improved organization of the device’s processes. (*Id.* ¶78.) It would also reflect the way voice assistants worked at the time. (*Id.* (citing Ex. 1015).) This approach would have been straightforward, as Shin’s device already integrates speech analysis and content generation. (*Id.* ¶79.) Thus, Shin also renders obvious these claim elements. (*Id.* ¶¶77-79.)

Even if Shin did not disclose or render obvious these units, they are disclosed by and would have been obvious in view of Aoyama. (*Id.* ¶¶80-84.) Aoyama discloses a “meaning analysis unit” that uses natural language processing to determine the “meaning” of the speech input and specify a “task” to be executed. (Ex. 1010 ¶¶[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, 4:49-58; Ex. 1002 ¶80.) 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) indicating what kind of information is wished to be acquired by the user.

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 ¶[0127]; *id.* ¶¶[0128]-[0130], [0174]-[0176], [0233], Figs. 3, 8, 13.) In the scheduling example, Aoyama discloses that the response component generation unit “acquires a search result” from a scheduling application. (*Id.* ¶[0129].) Thus, Aoyama discloses a notification-information acquiring unit configured to acquire notification information which includes content information (e.g., meeting title, participants) as a content information to be notified to the user based on the intention information (e.g., scheduling information). (Ex. 1002 ¶81.)

A POSITA would have been motivated to implement Aoyama’s “meaning analysis unit” and “response component generation unit” in Shin’s device for several reasons. (*Id.* ¶¶82-84.) First, Shin discloses a device that interprets and responds to voice inputs; a POSITA would have been motivated to look to other references, like Aoyama, to understand how speech input is processed, tasks are executed, and responses are generated. (*Id.* ¶82.) 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). (*Id.* ¶83); *KSR*, 550 U.S. at 417. 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. (*Id.*) 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. (*Id.*)

A POSITA 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. (*Id.* ¶84.)

Thus, Aoyama also renders obvious this claim element. (*Id.* ¶¶80-84.)

5. 1[e]: Output-Content Generating Unit

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.” In the ’675 patent, the “output-content generating unit” is the portion of the CPU that performs the recited function. (Ex. 1001, 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) and, consequently, this element cannot make the claim patentable. Regardless, Shin discloses this element. (Ex. 1002 ¶¶85-89.)

As discussed for claim elements 1[b]-1[d], Shin’s processor generates content for output to the user. (*Supra* §§VII.A.3-VII.A.4.) 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 ¶[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). ... Prepare your umbrella when you go out.” (*Id.* ¶[0088]; Ex. 1002 ¶87.) Text-to-speech technology converts the sentence to speech. (Ex. 1003 ¶[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. (Ex. 1002 ¶¶87.)

Shin’s output-content generating unit generates the output sentence based on notification information. As discussed, Shin’s processor analyzes the voice input to acquire notification information according to the user’s interests. (*Supra* §VII.A.4.) This notification information is incorporated into Shin’s output sentences, e.g., sky conditions, air temperature, and rainfall (in weather example) or current time (in time example). (Ex. 1003 ¶¶[0088], [0037]; Ex. 1002 ¶88.)

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. (Ex. 1002 ¶¶85-89.)

6. 1[f]: First Output Sentence

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.” For example, the ’675 patent’s device may respond to a spoken request about what “today’s schedule is.” (Ex. 1001, 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; *cf. id.*, 11:37-40; *id.*, Fig. 6 (“dinner” replaced with “meeting”).)

Shin and Aoyama together disclose this limitation. (Ex. 1002 ¶¶90-103.) Shin’s processor adjusts the amount of information provided based on the user-to-device distance. (Ex. 1003 ¶[0053]; *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 information decreases as user-to-device distance increases).) Shin’s Table 3 shows this:

TABLE 3

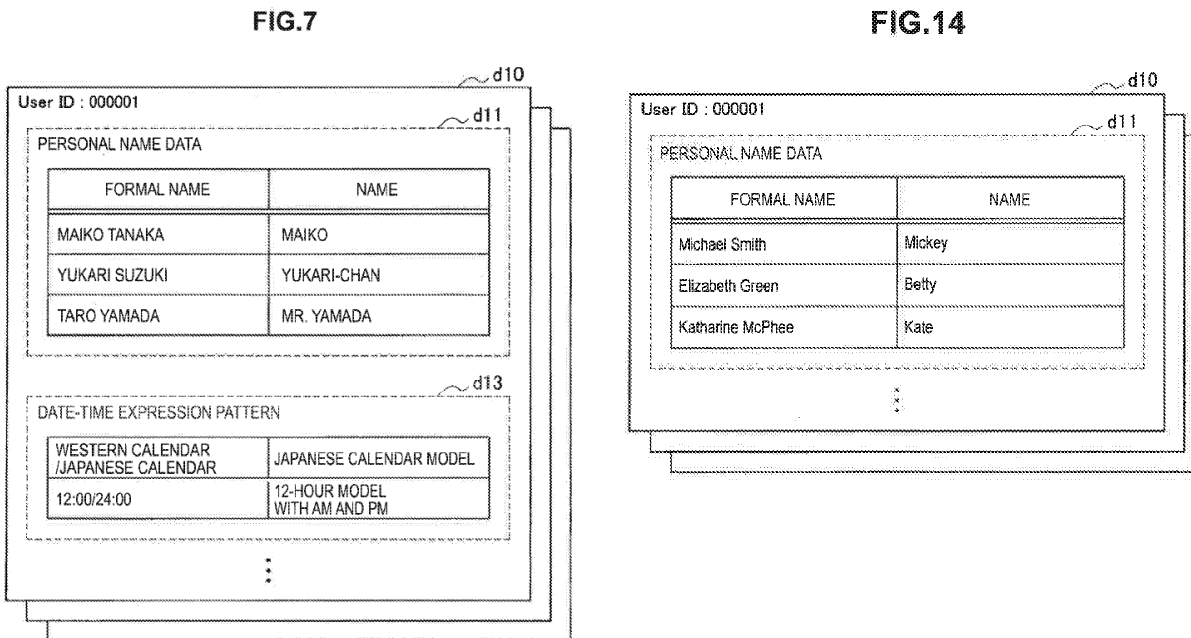
Distance (D) between user and electronic device	Output amount of information of content
$D < 1 \text{ m}$	Amount of information of original content
$1 \text{ m} \leq D < 2 \text{ m}$	Content of first abbreviation level
$2 \text{ m} \leq D < 3 \text{ m}$	Content of second abbreviation level
$3 \text{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

First Voice

(*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]; Ex. 1002 ¶92.) A POSITA would have readily understood Shin’s disclosure of abbreviating content to describe it more simply or provide the “gist” includes replacing words. (Ex. 1002 ¶92.)

Even if Shin did not expressly disclose replacing words to generate abbreviated/simplified content, it would have been obvious over Aoyama. (*Id.* ¶¶93-103.) Aoyama discloses a device (e.g., smartphone) that recognizes and analyzes a user’s speech, executes processing based on that input, and audibly presents the result as a sentence. (Ex. 1010 ¶¶[0042], [0051], [0127], [0161], [0241]-[0243], [0253], Fig. 8.) Aoyama’s device tailors output by replacing words based on the user or situation. (*Id.* ¶¶[0120]-[0124], [0138]-[0140], [0177], [0218].)

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



(*Id.*, Figs. 7, 14; *id.* ¶¶[0120]-[0124], [0234]-[0236]; Ex. 1002 ¶94.)

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 (Taro Yamada), honorific and surname (Mr. Yamada), honorific and full name (Mr. Taro Yamada), or first name only (Taro). (Ex. 1010 ¶¶[0050], [0126]; *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 their first name (Hirofumi) or familial role (Papa). (*Id.* ¶¶[0221], [0217]; *id.* ¶[0154] (“clear” replaces “nice”); Ex. 1002 ¶95.)

A POSITA would have been motivated to configure Shin’s processor to adjust the information provided based on the user-to-device distance by incorporating Aoyama’s word replacement feature for many reasons. (Ex. 1002 ¶¶96-103.)

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 to understand at greater distances. (*Id.* ¶97; Ex. 1010 ¶¶[0221], [0138]-[0140].)

Second, Aoyama teaches the desirability of replacing words based on user state (e.g., stopped, walking, running, laying, sitting, standing, speech speed, age, sex, etc.). (Ex. 1010 ¶¶[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 POSITA to use Shin’s proximity sensor to replace words based on the user’s distance, which a POSITA would have understood to be another example of user state. (Ex. 1002 ¶98.) Doing so also would have been consistent with Aoyama’s goal of making user-device interaction “more natural.” (Ex. 1010 ¶[0254].)

Third, it would have been obvious to combine Shin and Aoyama to address well-known privacy concerns, as recognized by Iyer. (Ex. 1002 ¶99.) Iyer discloses a device that processes and audibly responds to speech commands in discreet and normal modes. (Ex. 1012, 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 POSITA would have been motivated to incorporate Aoyama’s word replacement feature to replace words containing personal information with neutral alternatives. (Ex. 1002 ¶99 (citing Ex. 1011).)

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 (using word replacement to simplify content for distant users). (*Id.* ¶100); *KSR*, 550 U.S. at 417.

Fifth, such a configuration applies a known technique (Aoyama’s word replacement) to improve a similar device and method (Shin’s content tailoring system) in the same way. (*Id.*)

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 (providing content to distant users that is tailored using word replacement to address privacy concerns). (*Id.*)

Seventh, such a configuration would have been obvious to try. (Ex. 1002 ¶101); *KSR*, 550 U.S. at 417. Shin’s device presents only two design options for abbreviating or simplifying its output: omitting words or replacing them. (Ex. 1002 ¶101.) Given Shin’s disclosure of simplifying output to provide “core content” rather than “detailed content” for users farther away (Ex. 1003 ¶[0089]), it would have been obvious to try a word-replacement approach. (Ex. 1002 ¶101.)

A POSITA would have reasonably expected success in incorporating Aoyama’s word replacement feature into Shin’s processor for several reasons. (*Id.* ¶102.) First, this configuration would have been trivial, involving straightforward changes to enhance the processor’s approach to simplifying and abbreviating its output. (*Id.*) Second, such an approach was well known. For example, Ha (filed by Samsung) discloses a device that replaces “team meeting” and “date” (e.g., “detailed

information”) with “schedules” (e.g., “abstract information”) based on the user’s environment. (Ex. 1011 ¶¶[0097]-[0098], Fig. 7; *see generally id.*; Ex. 1002 ¶102.)

Thus, Shin and Aoyama render this limitation obvious. (Ex. 1002 ¶¶90-103.)

7. 1[g]: Second Output Sentence

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.”

The ’675 patent provides an example of this limitation using the scheduling example discussed above. (*Supra* §VII.A.6.) 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, 11:36-40.) In contrast with the example discussed above, here, “Mr. Yoshida” is not replaced with “Yoshi.” (*Supra* §VII.A.6.) 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, 11:26-37; Ex. 1002 ¶105.)

Shin discloses this claim element. (Ex. 1002 ¶¶104-08.) Shin discloses that, when a voice is classified as a second voice (e.g., <1 meter), the device generates a

second output comprising all available information. (Ex. 1003 ¶¶[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”), [0051].) This 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., <1 meter), the processor outputs unabbreviated content. (*Id.* ¶¶[0087]-[0088]; Ex. 1002 ¶106.)

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 ¶[0037] *with id.* ¶[0039] (shortened reply “nine ten”).) The word “time” conveys all intention information derived from the user statement, “Let me know what time it is now.” (Ex. 1002 ¶107.) The words “nine ten” represent all content information determined for that user statement. (*Id.*)

Thus, Shin discloses this limitation. (*Id.* ¶¶104-08.)

Accordingly, Shin and Aoyama render obvious claim 1. (*Id.* ¶¶49-108.)

B. Claim 2

Dependent claim 2 recites two additional elements, which Shin and Aoyama together disclose. (Ex. 1002 ¶¶109-17.)

1. 2[a]: Notification Information

Claim element 2[a] recites that “the notification information includes type information indicating a type of the content information.” Aoyama discloses this. (Ex. 1002 ¶¶110-13.)

The ’675 patent describes “type information” as information classifying the user’s desired content. (Ex. 1001, 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; Ex. 1002 ¶111.)

As discussed above, Aoyama’s device references replacement words stored in tables:

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, Fig. 7; *supra* §VII.A.6.) Aoyama discloses that its system includes a “response generation unit 17” that generates the device’s response to the user. (Ex. 1010 ¶[0124].) The response generation unit notifies another unit, “the response parameter generation unit 16 ... of the name corresponding to ‘Taro Yamada[.]’” (*Id.*) Thus, Aoyama’s system internally transmits the content information (e.g., “Taro Yamada”) along with type information (e.g., “name”). (*Id.*; Ex. 1002 ¶112.) The response parameter generation unit returns the replacement name (Mr. Yamada) from the appropriate table, determined using the type 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”).

Similarly, the titles of Aoyama’s tables are type information because they correspond to types of information the system looks up and replaces. (Ex. 1002 ¶113.)

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. (*Id.*; Ex. 1001, 14:41-43, 11:37-40.)

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

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.” For purposes of this Petition only, Petitioner assumes “placed” is intended to mean “replaced.”

The ’675 patent discloses that, in 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 can replace the words “Mr. Yoshida” with “Yoshi.” (Ex. 1001, 14:40-43, 11:37-40.) The word 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. (Ex. 1002 ¶¶114-16.)

As discussed, Aoyama discloses replacing names, including replacing “Taro Yamada” with “Mr. Yamada” according to the speech style of the user. (*Supra* §VII.A.6.) 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 ¶¶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.*; Ex. 1002 ¶116.) Thus, Aoyama discloses this limitation. (Ex. 1002 ¶¶114-16.)

Accordingly, Shin and Aoyama render obvious claim 2. (*Id.* ¶¶109-17.)

C. Claim 3

Claim 3 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.” 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, 14:23-27, 14:12-13, 14:36-43.) Aoyama discloses the same thing. (Ex. 1002 ¶¶118-22.)

Aoyama’s device references replacement words from tables when generating an output sentence. (*Supra* §§VII.A.6, VII.B.2.) For example, Aoyama’s device

reads the table titled “personal name data,” which stores “a ‘name’ corresponding to the ‘formal name’ [of] each person,” as shown in Figure 7:

FIG.7

The diagram shows a user profile card labeled 'd10' with 'User ID : 000001'. It contains two main sections: 'PERSONAL NAME DATA' (labeled 'd11') and 'DATE-TIME EXPRESSION PATTERN' (labeled 'd13'). The 'PERSONAL NAME DATA' section contains a table with two columns: 'FORMAL NAME' and 'NAME'. The 'DATE-TIME EXPRESSION PATTERN' section contains a table with two columns: 'WESTERN CALENDAR / JAPANESE CALENDAR' and 'JAPANESE CALENDAR MODEL'. Below the 'DATE-TIME EXPRESSION PATTERN' table, there are three vertical dots indicating more data.

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 ¶[0121], Fig. 7.) These tables, including their associations, are relationship information. (Ex. 1002 ¶120.) For example, the table shows that “Taro Yamada” is related to “Mr. Yamada.” (*Id.*) The table is stored in advance of generating the output sentence, so the words are predetermined. (Ex. 1010 ¶¶[0119], [0124]; *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. (Ex. 1002 ¶120.)

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, Aoyama replaces “Taro Yamada” with “Mr. Yamada” (*supra* §VII.A.6) to generate the sentence “You have a meeting with Mr. Yamada in room A,” where “Mr. Yamada” replaces “Taro Yamada.” (Ex. 1010 ¶¶[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. (Ex. 1002 ¶121.)

Accordingly, Shin and Aoyama render obvious claim 3. (*Id.* ¶¶118-22.)

D. Claim 4

Claim 4 depends from claim 3 and recites two additional elements, which Shin and Aoyama together disclose. (Ex. 1002 ¶¶123-30.)

1. 4[a]: Relationship Information

Claim element 4[a] recites that “the relationship information further includes information of type information indicating a type of the first word.” The ’675 patent’s contact list example (*see supra* §VII.C) describes that relationship information includes type information “person,” which indicates the type of the words “Mr. Yoshida.” (Ex. 1001, 14:30-36; Ex. 1002 ¶125.) Aoyama discloses this. (Ex. 1002 ¶¶124-26.)

Aoyama’s replacement word tables include type information. For example, the table in Aoyama titled “personal name data” includes “a ‘name’ corresponding to the ‘formal name’ [of] each person.” (Ex. 1010 ¶¶[0121], Fig. 7; *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. (Ex. 1002 ¶126.)

Thus, Aoyama discloses this limitation. (*Id.* ¶¶124-26.)

2. 4[b]: Output-Content Generating Unit

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.” In the ’675 patent’s example, the device selects the words “Mr. Yoshida” in the notification information to be replaced with “Yoshi.” (Ex. 1001, 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; Ex. 1002 ¶128.) Aoyama discloses this. (Ex. 1002 ¶¶127-29.)

As 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. (*Supra*

§VII.B.2.) Aoyama’s device determines this replacement by determining that the initial response includes the words “Taro Yamada,” which have a type “name.” (Ex. 1010 ¶[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 consults. (*Id.*; Ex. 1002 ¶129.) 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* Ex. 1010 ¶¶[0119], [0138], Figs. 7, 14.)

Accordingly, Shin and Aoyama render obvious claim 4. (*Id.* ¶¶123-30.)

E. Claim 5

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.” The phrase “the predetermined voice” lacks an antecedent basis; for purposes of this Petition only, Petitioner assumes “the

predetermined voice” refers to either “the first voice” or “the second voice” as recited in claim 1. Shin discloses this. (Ex. 1002 ¶¶131-38.)

Shin’s device can determine a user’s speech volume and adjust its output based, at least in part, on that volume. (Ex. 1003 ¶[0071]; *see also id.* ¶¶[0037]-[0039], [0049]-[0050], [0053]-[0054], [0076]-[0078], [0095], [0137], Abstract, claims 1-9, 15-17, 20; Ex. 1002 ¶132.) Shin discloses that its device detects a range of speech volumes and rates. (Ex. 1003 ¶¶[0037]-[0038].) Accordingly, a POSITA would have understood Shin’s disclosure to encompass receiving and processing as wide a range of speech as possible, including whispered speech. (Ex. 1002 ¶132.)

Even if Shin’s device did not explicitly support processing whispered speech, a POSITA would have been motivated for several reasons to configure the device to receive and process such inputs. (Ex. 1002 ¶¶133-37.)

First, the prior art suggests it. For example, Iyer teaches that the privacy concerns discussed above (*supra* §VII.A.6) may be addressed in part by processing whispered speech. (Ex. 1012, 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 POSITA to have Shin’s device process whispered speech to address privacy concerns. (Ex. 1002 ¶134.) Furthermore, Shin’s disclosure of processing speech at a wide range of volumes would have motivated a POSITA to process whispers, which a POSITA would have recognized are often quiet. (*Id.*)

Second, such a configuration would represent merely the addition of a known element (detecting 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). (*Id.* ¶135); *KSR*, 550 U.S. at 417. 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). (*Id.*) 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). (*Id.*)

Fifth, such a configuration would have been obvious to try. (Ex. 1002 ¶136); *KSR*, 550 U.S. at 417. As discussed, Shin’s device, which determines a user’s speech volume and adjusts its output accordingly (*e.g.*, Ex. 1003 ¶[0071]), demonstrates the foundational capability to process speech inputs of varying intensities and types. (*Id.* ¶[0037]-[0038]; Ex. 1002 ¶136.) Extending this functionality to include whispers would have been an obvious and predictable variation. (Ex. 1002 ¶136.)

A POSITA would have reasonably expected success in modifying Shin in this way for several reasons. (*Id.* ¶137.) First, Shin’s device already detects and responds to variations in speech volume. (Ex. 1003 ¶[0037]-[0038], [0071].) Extending its capabilities to handle whispers would involve straightforward adjustments (if

any). (Ex. 1002 ¶137.) Second, such functionality was known in the art. (*E.g.*, Ex. 1012, 3:61-64 (“the user may whisper voice commands to the microphone”); *infra* §XII; Ex. 1002 ¶137.)

Accordingly, Shin and Aoyama render obvious claim 5. (Ex. 1002 ¶¶131-38.)

F. Claims 6-7

Claim 6’s preamble recites “[a]n output-content control method.” Shin discloses both electronic devices and methods of operating them. (Ex. 1003 ¶¶[0181], [0003], [0008]-[0011], [0016]-[0021], [0043], [0047], [0092]-[0093], Title, Abstract, claims 15-19; Ex. 1002 ¶139.)

Claim 7’s preamble recites “[a] non-transitory storage medium that stores an output-content control program that causes a computer to execute” the same steps as in claim 6. Shin discloses a non-transitory storage medium. Specifically, Shin discloses devices implemented by instructions stored in a computer-readable storage media. (Ex. 1003 ¶[0181]; *id.* ¶¶[0011], [0179]-[0180], [0182], claim 20; Ex. 1002 ¶140.)

After the preamble, claims 6 and 7 are identical and merely recite the functions performed by the units in claim 1. For example, claims 6 and 7 recite “acquiring a voice” (claim 1’s voice acquiring unit, *supra* §VII.A.2), “calculating a distance” (performed in claim 1 by the proximity sensor, *supra* §VII.A.3.a), “analyzing the

acquired voice” (claim 1’s intention analyzing unit, *supra* §VII.A.4), “acquiring notification information” (claim 1’s notification-information acquiring unit, *supra* §VII.A.4), and “generating ... a first output sentence” and “generating ... a second output sentence” (claim 1’s output-content generating unit, *supra* §§VII.A.5-VII.A.7). As discussed above, Shin and Aoyama together disclose each of these functions and therefore render obvious claims 6 and 7. (*Supra* §VII.A; Ex. 1002 ¶141.)

VIII. GROUND 1B: CLAIMS 1-2 AND 5-7 WOULD HAVE BEEN OBVIOUS OVER SHIN AND IWASE.

Shin and Iwase render obvious claims 1-2 and 5-7. (Ex. 1002 ¶¶142-59.) Petitioner incorporates its discussion of Shin from Ground 1a. (*Supra* §VII.) To avoid repetition, Petitioner only addresses below the claims and limitations for which Iwase is relevant.

A. Claim 1

Iwase discloses, or at least suggests, the elements of claim 1 that recite replacing one word in the output content with another.³ (Ex. 1002 ¶¶143-52.)

Iwase discloses a device (e.g., smartphone) with a processor that audibly presents notification information to users. (Ex. 1017, 3:26-29, 5:2-7, 5:10-13, 5:63-6:3,

³ Because the claim language is included in Ground 1a, it is not repeated in subsequent Grounds.

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; Ex. 1002 ¶144.) 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.” (Ex. 1017, 15:6-14.)

A POSITA 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. (Ex. 1002 ¶¶145-51.)

First, Iwase suggests doing so. (*Id.* ¶146.) Iwase discloses that its word replacement system allows a user “to grasp notification contents more effectively.” (Ex. 1017, 2:11-14, 15:24-28, 21:66-22:8; Ex. 1002 ¶146.) A POSITA would have been motivated to incorporate Iwase’s word replacement feature into Shin’s device to gain the benefits disclosed by Iwase. (*Id.*)

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). (Ex. 1002 ¶147); *KSR*, 550 U.S. at 417.

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 (providing easier-to-grasp output). (*Id.*)

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 (delivering content to users at greater distances that is abbreviated using word replacement). (*Id.*)

Fifth, such a configuration would have been obvious to try. (Ex. 1002 ¶148); *KSR*, 550 U.S. at 417. Shin’s device presents only two design options for simplifying and abbreviating its output: removing words or replacing them. (Ex. 1002 ¶148.) Given Shin’s goal of providing “core content” rather than “detailed content” for distant users (Ex. 1003 ¶[0089]), it would have been obvious to try Iwase’s word replacement approach. (Ex. 1002 ¶148.)

Sixth, the similarity of the references’ outputs would have provided motivation. Shin discloses an output sentence: “Prepare your umbrella when you go out.” (Ex. 1003 ¶[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, 15:6-14.) A POSITA would have been motivated to modify Shin’s output, simplifying it for distant users by replacing words in “Prepare your umbrella when you go out” to yield: “Take an umbrella.” (Ex. 1002 ¶149.)

Lastly, the similarity of the devices would have motivated a POSITA 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, 9:59-10:9, 11:32-39, 12:4-9; Ex. 1002 ¶150.)

A POSITA would have reasonably expected success in incorporating Iwase’s word replacement into Shin’s processor for several reasons. (Ex. 1002 ¶151.) First, this configuration would have been trivial, involving straightforward changes to enhance the processor’s approach to simplifying and abbreviating its output. (*Id.*) Second, such an approach was well known. (*Id.* (citing Ex. 1011).) Third, 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. (*Id.* ¶151.)

Thus, Shin and Iwase render this limitation obvious. (*Id.* ¶¶143-52.) Because Shin discloses the remaining limitations of claim 1, Shin and Iwase render obvious claim 1. (*Id.*)

B. Claim 2

1. 2[a]: Notification Information

Iwase discloses this element. (Ex. 1002 ¶¶153-54.) 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, 14:35-44; *see also id.*, 14:24-49; Ex. 1002 ¶154.) For example, it may detect a clause

containing the phrase “I recommend.” (Ex. 1017, 14:44-45.) The device associates the semantic concept with the phrase in the notification information. (*Id.*, 14:35-40; Ex. 1002 ¶154.) The semantic concept is type information because it classifies the meaning of the associated phrase. (*Id.*) Thus, Iwase discloses this limitation. (Ex. 1002 ¶¶153-54.)

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

Iwase discloses this element. (*Id.* ¶¶155-58.) Iwase’s device determines the notification’s subject by identifying a phrase containing a semantic concept, which is type information. (*Supra* §VIII.B.1.) Once the subject is determined, words within the subject may be replaced based on user attributes. (Ex. 1017, 15:6-14; *supra* §VIII.A.) Consequently, 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). (Ex. 1002 ¶156.)

A POSITA would have been motivated to configure Shin’s processor to adjust the output information based on the user-to-device distance by applying Iwase’s word replacement feature, and would have reasonably expected success in doing so, for the reasons discussed above for claim 1. (*Supra* §VIII.A; Ex. 1002 ¶157.) A POSITA would have understood that Iwase’s word replacement feature would be applied at least to the information’s subject as Iwase discloses. (*Id.*)

Accordingly, Shin and Iwase render obvious claim 2. (Ex. 1002 ¶¶155-58.)

C. Claims 5-7

For the reasons discussed in Ground 1a and this Ground, Shin and Iwase render obvious claims 5-7. (*Supra* §§VII-VIII; Ex. 1002 ¶159.)

IX. GROUND 2A: CLAIMS 1-7 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA AND AOYAMA.

Shimomura and Aoyama render obvious claims 1-7. (Ex. 1002 ¶¶160-205.) Petitioner incorporates its discussion of Aoyama and how it discloses or renders obvious the relevant claim limitations discussed in Ground 1a. (*Supra* §VII.) To avoid repetition, Petitioner only addresses below the claims and limitations for which Shimomura is relevant.

A. Claim 1

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

Shimomura discloses a “speech control device that controls the speech of a device[.]” (Ex. 1004, claim 1; *id.*, Title.) Figure 1 shows an example:

[FIG. 1]

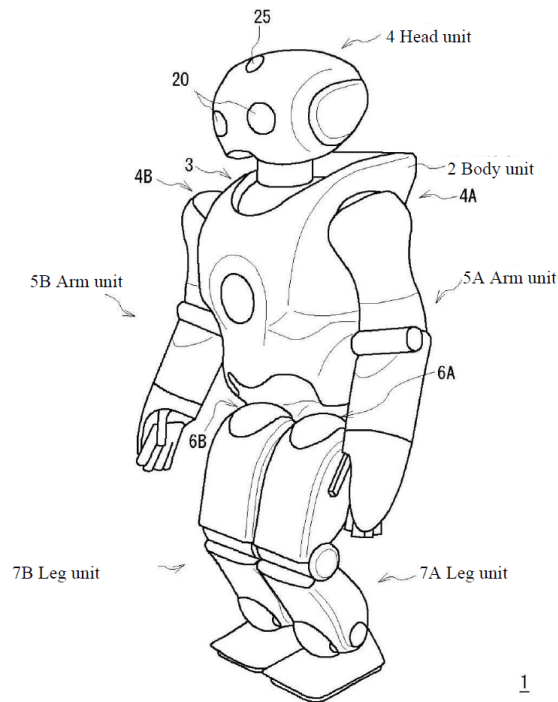


FIG. 1 Hardware constitution of the robot according to this embodiment (1)

(*Id.*, Fig. 1; Ex. 1002 ¶161.) The device receives speech via microphones, performs speech recognition, and determines the “content to be spoken[.]” (Ex. 1004 ¶[0046]; *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]; Ex. 1002 ¶162.) The device “controls speech forms ... according to the physical distance to the conversation partner.” (Ex. 1004 ¶¶[0049]-[0051]; *id.* ¶¶[0009]-[0010], [0016], [0057], [0085], [0088], [0090], [0092], claims 1-24.)

Thus, Shimomura discloses an “output-content control device.” (Ex. 1002 ¶¶161-63.)

2. 1[a]: Voice Acquiring Unit

Shimomura discloses a voice acquiring unit that is configured to “acquire a voice spoken by a user” as recited. (Ex. 1002 ¶¶164-65.) Shimomura discloses that its device, robot 1 (orange), includes a control unit 10 (blue) and a microphone 21 (red):

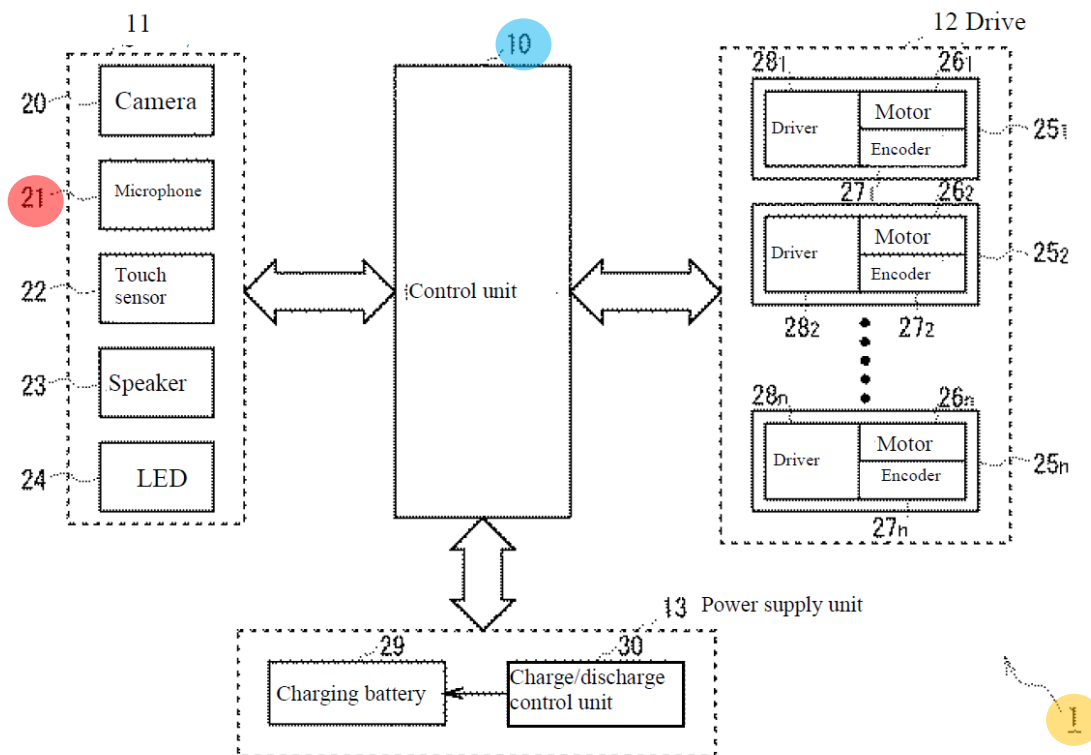


FIG. 2 Hardware constitution of the robot according to this embodiment (2)

(Ex. 1004, Fig. 2.) The control unit, which includes a CPU, controls overall operation and other data processing. (*Id.* ¶¶[0020], [0027], Fig. 3; *id.* ¶¶[0028], [0032], [0034].) The control unit includes a “speech recognition unit 42” that “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]; *id.* ¶¶[0036]-[0038], [0042], Fig. 4.)

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. (Ex. 1002 ¶¶164-65.)

3. 1[b]: Voice Classifying Unit

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

Shimomura’s device includes a “distance detection means for detecting the distance to the conversation partner user[.]” (Ex. 1004 ¶[0093].) Shimomura’s control unit includes an “image recognition unit.” (*Id.* ¶¶[0036]-[0038], [0042], Fig. 4.) That unit may detect and calculate a distance to the user “based on the image signal S1 from the CCD camera [sic].” (*Id.* ¶[0039]; *id.* ¶¶[0039] (“distance sensor”), [0059], [0070].)

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., cameras or other distance sensor). (Ex. 1002 ¶¶167-68.)

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

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. (Ex. 1002 ¶¶169-71.) Shimomura’s control unit includes a “speech synthesis unit” and a “speech character string transformation unit.” (Ex. 1004 ¶¶[0036]-

[0037], [0046], [0058], [0067], Fig. 4.) These units may “transform the character string” of the device’s spoken content “according to the distance to the conversation partner.” (*Id.* ¶¶[0058], [0067]-[0070]; *id.* ¶¶[0060]-[0063], [0073]-[0084], Abstract, claims 1-24.) To do so, Shimomura’s control unit first classifies the voice. (Ex. 1002 ¶169.)

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. (*Id.* ¶170.) 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

(*Id.*; Ex. 1004, Fig. 7, ¶¶[0060]-[0064].)

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. (Ex. 1002 ¶¶169-71.)

**4. 1[c]-1[d]: Intention Analyzing Unit and Notification-
Information Acquiring Unit**

Shimomura's control unit is configured to analyze the voice acquired by the voice acquiring unit. (Ex. 1004 ¶[0040]; Ex. 1002 ¶173.) It recognizes speech from the microphone's signal and, based on that speech, determines the output content. (*Supra* §§IX.A.1-IX.A.3.) Shimomura describes the user as a conversation partner (Ex. 1004 ¶¶[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 requested processing. (Ex. 1002 ¶173.)

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 ¶[0065]; Ex. 1002 ¶174.) Shimomura's control unit acquires from memory the required notification information, such as “Today's weather in Tokyo is sunny,” for output. (Ex. 1004 ¶¶[0044], [0065]; Ex. 1002 ¶174.)

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. (Ex. 1002 ¶175.) For example,

providing weather updates requires interpreting user intent and retrieving relevant content. (*Id.*) A POSITA 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.” (*Id.*) Thus, Shimomura expressly or inherently discloses these elements. (*Id.* ¶¶172-75.)

Even if Shimomura had not disclosed these elements, they would have been obvious from Shimomura. (*Id.* ¶¶176-78.) A POSITA 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.

For example, Shimomura describes voice processing capabilities that allow the device to interpret user intent and generate appropriate responses. A POSITA 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. (Ex. 1002 ¶¶177-78.) This modular design would facilitate consistent intent extraction, reliable content retrieval, and improved clarity and organization of internal processes. (*Id.*) It would also merely reflect the way known speech analysis systems worked at the time. (*Id.*; *see infra* §XIII.)

Thus, these claim elements would have been obvious over Shimomura. (*Id.* ¶¶176-78.)

Even if Shimomura had not disclosed or suggested these claim elements, they would not make the claim patentable because the '675 patent admits these claim elements were known. (*Supra* §VII.A.4.)

Lastly, and alternatively, these claim elements were disclosed by Aoyama (*id.*), and the claim as a whole would have been obvious in view of Shimomura and Aoyama. (Ex. 1002 ¶179.) A POSITA would have been motivated to implement the intention analyzing unit and notification information acquiring unit from Aoyama in Shimomura's device for the same reasons that a POSITA would have been motivated to similarly modify Shin. (*Id.*)

5. 1[e]: Output-Content Generating Unit

As discussed, Shimomura's control unit generates content for output. (*Supra* §§IX.A.3-IX.A.4.) The content can be a sentence including data like: "Today's weather in Tokyo is sunny." (Ex. 1004 ¶[0065]; *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 sentence is output as speech to the user. (*Id.* ¶[0047]; *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. (Ex. 1002 ¶181.)

Shimomura also discloses that the voice-content generating unit generates the output sentence based on the notification information. (*Id.* ¶182.) As discussed, Shimomura’s control unit determines the content based on the recognized speech so that the device converses with the user. (*Supra* §IX.A.4.) Thus, Shimomura discloses this element. (Ex. 1002 ¶¶180-82.)

6. 1[f]: First Output Sentence

Shimomura’s control unit implements the speech character string transformation unit to insert a word like “ne” after each phrase so the output “can be clearly understood according to the physical distance to the conversation partner.” (Ex. 1004 ¶¶[0057]-[0058]; *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]; Ex. 1002 ¶184.)

Thus, Shimomura discloses adjusting the output content based on the user-to-device distance, for example to increase user comprehension. (Ex. 1002 ¶¶183-85.) As discussed above, Aoyama discloses replacing words in output content, for example to adjust how names are presented. (*Supra* §§VII.A.6-VII.A.7.)

A POSITA would have been motivated to combine the disclosures of Shimomura and Aoyama to adjust Shimomura’s output using Aoyama’s word replacement feature for several reasons. (Ex. 1002 ¶¶186-91.)

First, Shimomura suggests doing so because it discloses that clear enunciation, particularly at distance, is advantageous. (*E.g.*, Ex. 1004 ¶¶[0004]-[0017]; Ex. 1002 ¶187.) A POSITA 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 POSITA to use the feature to advance Shimomura’s goals. (*Id.*)

Second, such a configuration represents the simple addition of one known element (Aoyama’s word replacement) to another known element (Shimomura’s dis-

tance-based output adjustments) to obtain predictable results (output content adjusted based on distance using word replacement). (*Id.* ¶188); *KSR*, 550 U.S. at 417. 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). (*Id.*) Fourth, such a configuration also applies a known technique (Aoyama’s word replacement) to a known device and method (Shimomura’s content adjustment) that is ready for improvement and yields predictable results (providing content, adjusted using word replacement, to users at greater distances). (*Id.*)

Fifth, a POSITA would have been motivated to implement the combination to address the privacy concerns discussed above. (*Supra* §VII.A.6; Ex. 1002 ¶189.) Such concerns would have motivated a POSITA to incorporate Aoyama’s word replacement feature into Shimomura’s device to replace words containing personal information with neutral alternatives. (*Id.*)

A POSITA would have reasonably expected success in incorporating Aoyama’s word replacement feature into Shimomura’s control unit for several reasons. (Ex. 1002 ¶190.) First, this configuration would have been trivial, involving straightforward changes to enhance the control unit’s approach to adjusting output. (*Id.*) Second, such an approach was well known. (*Id.*; *supra* §VII.A.6.)

Thus, Shimomura and Aoyama render this limitation obvious. (Ex. 1002 ¶¶183-91.)

7. 1[g]: Second Output Sentence

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 ¶¶[0062], [0092]; *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; Ex. 1002 ¶193.)

For example, Shimomura discloses that the output sentence when the voice is classified as the second voice may be: “Today’s weather in Tokyo is sunny.” (Ex.

1004 ¶[0065].) In this example, the word “weather” conveys all intention information derived from a user statement like “Let me know today’s weather.” (Ex. 1002 ¶194.) The words “Tokyo” and “sunny” represent all content information determined for that user statement. (*Id.*)

Thus, Shimomura discloses this element. (*Id.* ¶¶192-94.)

Accordingly, Shimomura and Aoyama render obvious claim 1. (*Id.* ¶¶161-95.)

B. Claims 2-4

Aoyama discloses the additional limitations of claims 2-4. (*Supra* §§VII.B-VII.D.) For the reasons discussed in Ground 1a and this Ground, Shimomura and Aoyama render obvious claims 2-4. (Ex. 1002 ¶196.)

C. Claim 5

Shimomura discloses, or at least renders obvious, the additional limitation of claim 5. (Ex. 1002 ¶¶197-203.) Shimomura discloses that its device receives and processes speech input from a user. (*Supra* §IX.A.2.) A whisper is a form of speech. (Ex. 1002 ¶198.) Therefore, Shimomura discloses classifying a whispered voice. (*Id.*)

Even if Shimomura did not teach receiving and processing whispered speech, a POSITA would have been motivated for several reasons to configure the device to do so. (*Id.* ¶¶199-203.)

First, the prior art suggests it. For example, Iyer teaches that the privacy concerns discussed above (*supra* §VII.A.6) may be addressed in part by processing whispered speech, motivating a POSITA to have Shimomura’s device do so. (Ex. 1012, 3:61-64; Ex. 1002 ¶200.) Furthermore, Shimomura’s general disclosure of processing speech (*supra* §IX.A) would have motivated a POSITA to process as wide a range of speech as possible, including whispers. (Ex. 1002 ¶200.)

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). (*Id.* ¶201); *KSR*, 550 U.S. at 417. 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). (*Id.*) Fourth, such a configuration reflects applying a known technique (whisper processing) to achieve predictable results (content adjustment based on whispered input). (*Id.*)

A POSITA would have reasonably expected success in modifying Shimomura in this way for several reasons. (Ex. 1002 ¶202.) First, extending Shimomura’s capabilities to handle whispers would involve straightforward adjustments, if any. (*Id.*) Second, such functionality was known in the art. (*Id.* (citing Exs. 1012, 1016).)

Accordingly, Shimomura and Aoyama render obvious claim 5. (*Id.* ¶¶197-203.)

D. Claims 6-7

Shimomura discloses both a speech control device and method. (Ex. 1004 ¶¶[0012]-[0013]; *id.* ¶¶[0001], [0009], [0016], [0057], Title, Abstract, Figs. 6, 8, 13, claims 9-16; Ex. 1002 ¶204.) Shimomura's devices are implemented by instructions stored in a computer-readable storage media. (Ex. 1004 ¶¶[0027], [0030], [0036]; *id.* ¶¶[0028]-[0029], [0037]-[0047], [0051]-[0083], Figs. 3-4; Ex. 1002 ¶204.)

As discussed, the remainder of claims 6-7 are identical and recite the functions performed by the units in claim 1. (*Supra* §VII.F; Ex. 1002 ¶205.) Thus, for the reasons provided for claim 1, Shimomura and Aoyama render obvious claims 6 and 7. (*Id.*)

X. GROUND 2B: CLAIMS 1-2 AND 5-7 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA AND IWASE.

Shimomura and Iwase render obvious claims 1-2 and 5-7. (Ex. 1002 ¶¶206-15.) Petitioner incorporates its discussion of Shimomura (Ground 2a) and Iwase (Ground 1b) and how they disclose or render obvious the relevant claim limitations. (*Supra* §§IX, VIII.)

A. Claim 1

Shimomura and Iwase together disclose each limitation of claim 1. (*Supra* §§VIII.A, IX.A; Ex. 1002 ¶¶207-12.)

A POSITA would have been motivated to configure Shimomura’s control unit to adjust the output 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. (Ex. 1002 ¶¶208-12.)

First, Iwase suggests doing so. (*Id.* ¶209.) Iwase’s system allows a user “to grasp notification contents more effectively.” (Ex. 1017, 2:11-14, 15:24-28, 21:66-22:8.) Thus, a POSITA would have been motivated to incorporate Iwase’s word replacement feature into Shimomura’s device to make the output content easier to understand. (*Id.*; Ex. 1002 ¶209)

Second, such a configuration represents 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 adjusted using word replacement). (Ex. 1002 ¶210); *KSR*, 550 U.S. at 417.

Third, such a configuration reflects applying a known technique (Iwase’s word replacement) to improve a similar device and method (Shimomura’s content adjustment) in the same way. (*Id.*)

Fourth, such a configuration also applies a known technique (Iwase’s word replacement) to a known device and method (Shimomura’s) that is ready for improvement and yields predictable results (providing content adjusted via word replacement to users at greater distances). (*Id.*)

A POSITA would have reasonably expected success in incorporating Iwase's word replacement feature into Shimomura's control unit for several reasons. (Ex. 1002 ¶211.) First, this configuration would have been trivial, involving straightforward changes to enhance the control unit's approach to adjusting its output content. (*Id.*) Second, such an approach was well known. (*Id.* (citing Ex. 1011).)

Accordingly, Shimomura and Iwase render obvious claim 1. (Ex. 1002 ¶¶207-12.)

B. Claim 2

Iwase discloses the additional limitations of claim 2. (*Supra* §VIII.B.) A POSITA would have been motivated to configure Shimomura's control unit to adjust the information provided 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 discussed above for claim 1. (Ex. 1002 ¶213.)

Accordingly, Shimomura and Iwase render obvious claim 2. (*Id.* ¶¶213-14.)

C. Claims 5-7

For the reasons discussed in Grounds 1a-2a and this Ground, Shimomura and Iwase render obvious claims 5-7. (*Supra* §§VII-IX; Ex. 1002 ¶215.)

XI. GROUND 3: CLAIMS 3-4 WOULD HAVE BEEN OBVIOUS OVER THE COMBINATIONS IN GROUNDS 1B OR 2B AND AOYAMA.

Ground 1b (Shin + Iwase) and Ground 2b (Shimomura + Iwase) show how those references render obvious claims 1-2 and 5-7. The additional limitations of claims 3 and 4 are disclosed by Aoyama, and thus would have been obvious in view of the references in Ground 1b and Aoyama and in view of the references in Ground 2b and Aoyama. (Ex. 1002 ¶¶216-24.)

A. Claim 3

Aoyama discloses the additional limitations of claim 3. (*Supra* §§VII.A.6, VII.B.2, VII.C; Ex. 1002 ¶¶217-18.) 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. (Ex. 1002 ¶¶219-22.) For example, a POSITA 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 Iwase, such as “request,” “proposal,” “wish,” or “opinion,” for several reasons. (*Id.* ¶219.)

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 re-

sults (generating responses by looking up words in semantic concept tables and replacing those words with more general ones from the tables). (*Id.* ¶220); *KSR*, 550 U.S. at 417. 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. (*Id.*) 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). (*Id.*)

A POSITA 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. (Ex. 1002 ¶221.) This configuration would have been trivial, involving straightforward changes to enhance the processor’s approach to simplifying and abbreviating its output content. (*Id.*)

Accordingly, claim 3 would have been obvious over (a) Shin, Iwase, and Aoyama and (b) Shimomura, Iwase, and Aoyama. (*Id.* ¶¶217-22.)

B. Claim 4

A POSITA 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. (*Supra* §XI.A; Ex. 1002 ¶223.) 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”). (*Id.*) 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. (*Supra* §§VII-VIII.) Accordingly, either Shin or Shimomura, in combination with Iwase and Aoyama, render obvious claim 4. (Ex. 1002 ¶¶223-24.)

XII. GROUND 4: CLAIM 5 WOULD HAVE BEEN OBVIOUS OVER THE COMBINATIONS IN GROUNDS 1A-2B AND SCHUSTER.

Claim 5 depends from claim 1 and would have been obvious in view of the references in Grounds 1a-2b, in further view of Schuster. (*Id.* ¶¶225-31.)

Schuster discloses a device that receives verbal queries and provides voice outputs like search answers or recommendations. (Ex. 1016, 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 micro-

phone signal analysis (Ex. 1016, 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. (Ex. 1002 ¶227.) Thus, even if PO argues that claim 5 requires the voice to be classified based on whether it is a whisper, the claim would have been obvious.

A POSITA 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. (*Id.* ¶¶228-31.)

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). (*Id.* ¶229); *KSR*, 550 U.S. at 417. 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. (*Id.*) 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). (*Id.*) Fourth, Schuster discloses classifying based on distance (Ex. 1016, 3:27-30), just as Shin and Shimomura do.

A POSITA would have reasonably expected success in modifying Shin and/or Shimomura in this way. (Ex. 1002 ¶230.) Shin and Shimomura’s devices already classify speech into two or more categories. (*Id.*; *see supra* §§VII, IX.) 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. (Ex. 1002 ¶230.)

Accordingly, the references in Grounds 1a-2b and Schuster render obvious claim 5. (*Id.* ¶¶225-31.)

**XIII. GROUND 5: CLAIMS 1-7 WOULD HAVE BEEN
OBVIOUS OVER THE COMBINATIONS IN GROUNDS
1A-4 AND KRISTJANSSON.**

As discussed above, the references in Grounds 1a-4 disclose each limitation of claims 1-7. (*Supra* §§VII-XII.) The claims also would have been obvious over those references and Kristjansson. Specifically, if PO argues that Shin and/or Shimomura do 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. (Ex. 1002

¶¶232-39.) These limitations refer to determining the user’s intent using natural language understanding (“NLU”) processing and acquiring the desired information. (Ex. 1001, 4:25-27; Ex. 1002 ¶233.)

Kristjansson discloses a speech-controlled device. (Ex. 1013, 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.*) To do so, Kristjansson’s system includes “an NLU component” that analyzes the user’s speech and “derive[s] an intent or a desired action from the user[.]” (*Id.*, 7:26-67.) The intent information is sent to a “command processor” that retrieves content based on the intent. (*Id.*, 10:38-11:12.) The retrieved content is provided to the user. (*Id.*, 10:38-11:12.) Thus, Kristjansson discloses elements 1[c]-1[d], 6[c]-6[d] and 7[c]-7[d]. (Ex. 1002 ¶234.)

As discussed, Shin and Shimomura, like Kristjansson, disclose devices that receive voice input and return voice output. (*Supra* §§VII, IX.) A POSITA would have been motivated to implement Kristjansson’s NLU teachings in Shin and/or Shimomura’s devices for several reasons. (Ex. 1002 ¶¶235-39.)

First, Shin and Shimomura suggest doing so. (*Id.* ¶236.) Shin and Shimomura disclose devices that analyze voice input, identify content of interest, and generate the requested content. (*Supra* §§VII, IX.) Kristjansson provides additional detail about how to structure such a device, with NLU processing and content retrieval.

(Ex. 1002 ¶236.) A POSITA would have looked to Kristjansson when implementing Shin and/or Shimomura's devices. (*Id.*)

Second, Kristjansson's disclosure represents a routine and well-established approach to providing a user with requested content. (*Id.* ¶237; *see* Ex. 1015, 431-33); *KSR*, 550 U.S. at 417. Third, such a modification applies a known technique (Kristjansson's NLU processing and information retrieval) to achieve predictable results (providing user with content). (*Id.*) Fourth, such a modification reflects 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 requested content). (*Id.*)

A POSITA 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, 431-33; Ex. 1002 ¶238) and because Shin and Shimomura contemplate modification or addition to their systems (*e.g.*, Ex. 1003 ¶[0184]; Ex. 1004 ¶[0089]). (Ex. 1002 ¶238.) 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, claim 1, 3:1-27; Ex. 1002 ¶238.)

Accordingly, the Grounds 1a-4 references, in further view of Kristjansson, render obvious claims 1-7. (Ex. 1002 ¶¶232-39.)

XIV. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

Secondary considerations should be considered but do not control the obviousness conclusion. *Newell Cos. v. Kenney Mfg. Co.*, 864 F.2d 757, 768 (Fed. Cir. 1988). Where, as here, a strong *prima facie* obviousness showing exists, even relevant secondary considerations supported by substantial evidence may not dislodge the obviousness conclusion. *Leapfrog Enters. v. Fisher-Price, Inc.*, 485 F.3d 1157, 1162 (Fed. Cir. 2007). Petitioner is aware of no evidence supporting a claim for secondary considerations.

XV. DISCRETIONARY DENIAL UNDER §314(A) IS NOT APPROPRIATE.

The Board should not exercise its discretion under §314(a). PO asserted the '675 patent against Petitioner in district court. *SoundClear Techs. LLC v. Amazon.com, Inc. et al.*, No. 1:24-cv-01283 (AJT/WBP) (E.D. Va.) (“the Litigation”). On November 8, 2024, the court granted Petitioner’s motion to dismiss, held all claims invalid under §101, and entered judgment in Petitioner’s favor. (Ex. 1008.) On January 10, 2025, the Court granted PO’s motion to alter the judgment, finding that “a plausible claim for patent validity exists” in view of PO’s new arguments. (Ex. 1014.) As of this Petition’s filing, the Court has not entered a schedule or set a trial date. Regardless of when a trial may occur, discretionary denial is not warranted for the reasons below.

A. Petitioner’s *Sotera* Stipulation

The Board “will not discretionarily deny institution in view of parallel district court litigation where a petitioner presents a stipulation not to pursue in a parallel proceeding the same grounds or any grounds that could have reasonably been raised” in the IPR petition. Katherine K. Vidal, *Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation* (June 21, 2022) (“Director Memo”) at 3, 7. Petitioner so stipulates here. This stipulation is sufficient to avoid denial under §314(a). *Sotera Wireless, Inc. v. Masimo Corp.*, IPR2020-01019, Paper 12 (P.T.A.B. Dec. 1, 2020); *Endymed Med. Ltd. v. Serendia, LLC*, IPR2024-00843, Paper 14 at 13 (P.T.A.B. Jan. 10, 2025).

B. Compelling Evidence of Unpatentability

The Board “will not ... discretionarily deny institution in view of parallel district court litigation where a petition presents compelling evidence of unpatentability.” (Director Memo at 4.) “Compelling, meritorious challenges are those in which the evidence, if unrebutted in trial, would plainly lead to a conclusion that one or more claims are unpatentable by a preponderance of the evidence.” (*Id.* at 4.) That is the case here, as the references relied on plainly render obvious the challenged claims.

Fairness also warrants institution, as denial would forever time-bar Petitioner from seeking IPR in the event that the district court case were to later terminate based

on a finding of noninfringement. The public’s interest in ensuring patent validity also supports institution.

XVI. DISCRETIONARY DENIAL UNDER §325(D) IS NOT APPROPRIATE.

Both prongs of the *Advanced Bionics* framework favor institution. *Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH*, IPR2019-01469, Paper 6 at 8 (P.T.A.B. Feb. 13, 2020) (precedential).

Prong 1. Shin, Shimomura, Iwase, and Kristjansson were never before the Examiner. At least one of these references is relied on for each challenged claim in each Ground herein. This is sufficient to avoid denial under §325(d). *Shenzen Chic Elecs. v. Pilot, Inc.*, IPR2023-00810, Paper 12 at 21 n.11 (P.T.A.B. Nov. 8, 2023) (denial inappropriate where challenges based on new art/arguments “address all challenged claims,” regardless of whether other challenges relied on previously considered art); *TP-Link Corp. Ltd. v. Netgear, Inc.*, IPR2023-01469, Paper 10 at 9-10 (P.T.A.B. Apr. 2, 2024). Because the first prong is unmet, the Board “need not address the second” prong. *JUUL Labs, Inc. v. NJOY, LLC*, IPR2024-00160, Paper 10 at 13 (P.T.A.B. May 24, 2024); *Quasar Sci. LLC v. Colt Int’l Clothing, Inc.*, IPR2023-00611, Paper 10 at 12 (P.T.A.B. Oct. 10, 2023).

Prong 2. Should the Board reach the second prong, it too favors institution. The Examiner did not consider Shin, Shimomura, Iwase, or Kristjansson. Nor did the Examiner benefit from this Petition’s mapping and accompanying expert

declaration. Where, as here, “the Office has not substantively addressed prior art ... the petitioner may show material error by showing that the Office failed to reject the claims over the asserted art.” *Quasar*, Paper 10 at 14; *Advanced Bionics*, Paper 6 at 10 (“[A] petitioner may show the Office erred by overlooking something persuasive.”); *Apple Inc. v. Telefonaktiebolaget LM Ericsson*, IPR2022-00457, Paper 7 at 8 (P.T.A.B. Sept. 21, 2022). The disclosures presented here show that the Examiner erred in allowing the claims.

XVII. MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT

Pursuant to 37 C.F.R. §42.8(a)(1), the mandatory notices identified in 37 C.F.R. § 42.8(b) are provided below as part of this petition.

A. Real Party-In-Interest (37 C.F.R. §42.8(b)(1))

Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. are the real parties-in-interest.

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

To the best knowledge of Petitioner, the ’675 patent is or has been involved in the following district court litigations:

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

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

Petitioner provides the following designation of counsel, all of whom are included in Customer No. 20,995 identified in Petitioner’s Power of Attorney.

Lead Counsel	Back-up Counsel
Colin B. Heideman (Reg. No. 61,513) 2cbh@knobbe.com BoxSEAZN2L2104LP3@knobbe.com	Joseph R. Re (Reg. No. 31,291) 2jrr@knobbe.com
<u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson, & Bear, LLP 925 4th Ave., Ste. 2500 Seattle, WA 98104 Telephone: 206-405-2000 Facsimile: 206-405-2001	<u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson, & Bear, LLP 2040 Main Street, 14th Floor Irvine, CA 92614 Telephone: (949) 760-0404 Facsimile: (949) 760-9502
	Nathan D. Reeves (Reg. No. 77,806) 2ndr@knobbe.com Logan P. Young (Reg. No. 79,294) 2lpy@knobbe.com
	<u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson & Bear, LLP 925 4th Ave., Ste. 2500 Seattle, WA 98104 Telephone: (206) 405-2000 Facsimile: (206) 405-2001

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

Please direct all correspondence to lead counsel and back-up counsel at the addresses shown above. Petitioner also consents to electronic service by email to BoxSEAZN2L2104LP3@knobbe.com.

E. Grounds for Standing (37 C.F.R. §42.104)

Petitioner certifies that the '675 patent is available for IPR and is not barred or estopped from requesting IPR on the identified grounds.

F. Payment of Fees (37 C.F.R. §42.15(a))

The Office may charge the §42.15(a) fee to Deposit Account No. 11-1410. Review of seven claims is requested. Payment for any additional fees due may be charged to the above-referenced Deposit Account.

Dated: February 27, 2025 By: /Colin B. Heideman/
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP
Attorney for Petitioners Amazon.com, Inc.,
Amazon.com Services LLC, and Amazon Web
Services, Inc.

APPENDIX

Listing of Claims from U.S. 11,244,675	
Claim 1	
1[pre]	An output-content control device comprising:
1[a]	a voice acquiring unit configured to acquire a voice spoken by a user;
1[b]	a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;
1[c]	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;
1[d]	a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and
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, wherein
1[f]	the output-content generating unit is further configured
	to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word, and
1[g]	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.

Claim 2	
2[pre]	The output-content control device according to claim 1, wherein
2[a]	the notification information includes type information indicating a type of the content information, and
2[b]	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.
Claim 3	
--	The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.
Claim 4	
4[pre]	The output-content control device according to claim 3, wherein
4[a]	the relationship information further includes information of type information indicating a type of the first word, and
4[b]	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.

Claim 5	
--	The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.
Claim 6	
6[pre]	An output-content control method comprising:
6[a]	acquiring a voice spoken by a user;
6[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;
6[c]	analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;
6[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
6[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
6[f]	generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.

Claim 7	
7[pre]	A non-transitory storage medium that stores an output-content control program that causes a computer to execute:
7[a]	acquiring a voice spoken by a user;
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;
7[c]	analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;
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
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
7[f]	generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that this **PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 11,244,675** contains 13,671 words according to the word-processing program used to prepare this paper. The foregoing word count complies with the 14,000 word type-volume limit specified by 37 C.F.R. § 42.24(a)(1).

Dated: February 27, 2025 By: / Colin B. Heideman/
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on the date below a copy of this **PE-**
TITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 11,244,675
and **EXHIBITS 1001-1018**, are being served on February 27, 2025 via Federal
Express overnight mail on counsel of record for U.S. Patent No. 11,244,675, at the
Correspondence Address of record below:

Chandran Iyer
Daignault Iyer LLP
8229 Boone Blvd., Suite 450
Vienna, VA 22182

A courtesy copy is also being served via email on counsel for the patent holder
in the pending litigation:

Daignault Iyer LLP
8229 Boone Blvd., Suite 450
Vienna, VA 22182
Email: DI-SoundClear@daignaultiyer.com

Dated: February 27, 2025 By: /Colin B. Heideman/
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP