

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. IPR2025-00565
Patent No. 11,069,337

**PETITION FOR *INTER PARTES* REVIEW
OF CLAIMS 1-5 OF U.S. PATENT NO. 11,069,337**

TABLE OF CONTENTS

I.	INTRODUCTION -----	1
II.	BACKGROUND -----	2
III.	THE '337 PATENT -----	3
	A. Overview -----	3
	B. Prosecution History -----	5
IV.	RELIEF REQUESTED -----	6
	A. Grounds -----	6
	B. The References Are Analogous Prior Art -----	6
V.	LEVEL OF ORDINARY SKILL -----	7
VI.	CLAIM CONSTRUCTION -----	8
VII.	GROUND 1: CLAIMS 1-5 ARE ANTICIPATED BY SHIN -----	8
	A. Claim 1 -----	8
	1. 1[pre]: Voice-Content Control Device -----	8
	2. 1[a]: Proximity Sensor -----	9
	3. 1[b]: Voice Classifying Unit -----	11
	4. 1[c]: Process Executing Unit -----	15
	5. 1[d]: Voice-Content Generating Unit -----	16
	6. 1[e]: Output Controller -----	18
	7. 1[f]: Generate a First Output Sentence -----	18
	8. 1[g]: Generate a Second Output Sentence -----	20
	9. 1[h]: Output Controller Adjusts Volume of Voice Data -----	22

B.	Claim 2 -----	24
1.	2[a]: Process Executing Unit-----	24
2.	2[b]: Voice-Content Generating Unit-----	27
C.	Claim 3 -----	27
D.	Claims 4 and 5-----	29
VIII.	GROUND 2: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIN-----	30
A.	Claim 1 -----	30
B.	Claim 2 -----	32
C.	Claim 3 -----	33
D.	Claims 4 and 5-----	35
IX.	GROUND 3: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA -----	35
A.	Claim 1 -----	35
1.	1[pre]: Voice-Content Control Device -----	35
2.	1[a]: Proximity Sensor-----	37
3.	1[b]: Voice Classifying Unit-----	37
4.	1[c]: Process Executing Unit-----	40
5.	1[d]: Voice-Content Generating Unit-----	41
6.	1[e]: Output Controller -----	42
7.	1[f]: Generate a First Output Sentence -----	43
8.	1[g]: Generate a Second Output Sentence -----	44
9.	1[h]: Output Controller Adjusts Volume of Voice Data -----	46

B.	Claim 2 -----	53
1.	2[a]: Process Executing Unit-----	53
2.	2[b]: Voice-Content Generating Unit-----	56
C.	Claim 3 -----	56
D.	Claims 4 and 5-----	59
X.	GROUND 4: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA AND SHIN -----	59
A.	Claim 1 -----	59
B.	Claim 2 -----	62
C.	Claim 3 -----	62
D.	Claims 4 and 5-----	64
XI.	GROUND 5: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIN AND/OR SHIMOMURA IN VIEW OF KRISTJANSSON-----	65
XII.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	68
XIII.	DISCRETIONARY DENIAL UNDER §314(A) IS NOT APPROPRIATE -----	68
A.	Petitioner’s <i>Sotera</i> Stipulation-----	69
B.	The Petition Presents Compelling Evidence of Unpatentability -----	69
XIV.	DISCRETIONARY DENIAL UNDER §325(D) IS NOT APPROPRIATE -----	70
XV.	MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT -----	72
A.	Real Party-In-Interest (37 C.F.R. §42.8(b)(1))-----	72
B.	Related Matters (37 C.F.R. §42.8(b)(2))-----	72

C.	Lead and Backup Counsel (37 C.F.R. §42.8(b)(3)) -----	73
D.	Service Information (37 C.F.R. §42.8(b)(4)) -----	74
E.	Grounds for Standing (37 C.F.R. §42.104) -----	74
F.	Payment of Fees (37 C.F.R. §42.15(a)) -----	74

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) -----	70, 72
<i>Apple Inc. v. Telefonaktiebolaget LM Ericsson,</i> IPR2022-00457, Paper 7 (P.T.A.B. Sept. 21, 2022) -----	72
<i>Endymed Med. Ltd. v. Serendia, LLC,</i> IPR2024-00843, Paper 14 (P.T.A.B. Jan. 10, 2025) -----	69
<i>In re GPAC Inc.,</i> 57 F.3d 1573 (Fed. Cir. 1995)-----	7
<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)-----	<i>passim</i>
<i>Leapfrog Enters. v. Fisher-Price, Inc.,</i> 485 F.3d 1157 (Fed. Cir. 2007) -----	68
<i>Newell Cos. v. Kenney Mfg. Co.,</i> 864 F.2d 757 (Fed. Cir. 1988)-----	68
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.,</i> 868 F.3d 1013 (Fed. Cir. 2017) -----	8
<i>Quasar Sci. LLC v. Colt Int’l Clothing, Inc.,</i> IPR2023-00611, Paper 10 (P.T.A.B. Oct. 10, 2023)-----	71, 72
<i>Samsung Elecs. Co. v. Maxell, Ltd.,</i> IPR2024-00867, Paper 9 (P.T.A.B. Nov. 7, 2024) -----	71
<i>Shenzen Chic Elecs. v. Pilot, Inc.,</i> IPR2023-00810, Paper 12 (P.T.A.B. Nov. 8, 2023)-----	70
<i>Sotera Wireless, Inc. v. Masimo Corp.,</i> IPR2020-01019, Paper 12 (P.T.A.B. Dec. 1, 2020)-----	69

<i>TP-Link Corp. Ltd. v. Netgear, Inc.</i> , IPR2023-01469, Paper 10 (P.T.A.B. Apr. 2, 2024) -----	70
---	----

<i>Unwired Planet, LLC v. Google Inc.</i> , 841 F.3d 995 (Fed. Cir. 2016)-----	7
---	---

<i>Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.</i> , 200 F.3d 795 (Fed. Cir. 1999)-----	8
---	---

Statutes and Rules:

35 U.S.C. § 102 -----	6, 7
-----------------------	------

35 U.S.C. § 103 -----	6
-----------------------	---

35 U.S.C. § 314 -----	68, 69
-----------------------	--------

35 U.S.C. § 325 -----	70, 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) -----	69
--	----

LIST OF EXHIBITS

No.	Description
1001	U.S. Patent No. 11,069,337
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. 11,069,337
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 <i>Fundamenta Informaticae</i> 425 (2014) (“Duta”)

Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively, “Petitioner” or “Amazon”) request *inter partes* review of claims 1-5 of U.S. Patent No. 11,069,337 (“the ’337 patent”), which SoundClear Technologies LLC (“Patent Owner” or “PO”) purportedly owns.

I. INTRODUCTION

The ’337 patent describes an electronic device, such as a speaker or phone, that receives and processes a user’s voice and outputs a voice response. (Ex. 1001 (’337 patent), Fig. 1, 2:62-3:1.) The patent claims priority to 2018, years after Apple launched Siri (2011), Amazon launched Alexa-enabled devices (2014), and Google launched its voice assistant (2016). The patent admits that devices that detect voice, perform processing according to the user’s intent, and provide a voice output were known. (*Id.*, 1:21-28.)

The Examiner allowed the ’337 patent claims because they recite calculating the distance between the device and user, then adjusting the content and volume of the response based on that distance. (Ex. 1007, 28-32, 41-46.) But this was not new in 2018. It had been disclosed in many prior art references, including the references relied on herein.

Because the ’337 patent removes from the public store of knowledge devices that were known and obvious to those skilled in the art, the Board should cancel the claims.

II. BACKGROUND

Using a proximity sensor to determine a user’s distance from a device and then tailoring the information provided and the volume based on that distance has been known for decades. (Ex. 1002 ¶¶35-44.) In 2005, Shimomura (filed by Sony) described such a device. (Ex. 1004, Abstract.) It used cameras to determine the distance between the device and the user (*id.* ¶¶[0039], [0093]), classified the user’s speech based on distance (e.g., whether the user was more or less than 350 cm away) and adjusted the output content (e.g., including or omitting words) and volume accordingly (*id.* ¶¶[0049]-[0051], [0057], [0086]-[0088]). (Ex. 1002 ¶41.)

In 2017, Shin (filed by Samsung) disclosed a device that acquires a user’s voice input via a microphone, analyzes it, and provides voice output via a speaker. (Ex. 1003 ¶¶[0031], [0037].) Shin used a “distance detection module” or “proximity sensor” to “compute a distance between a user and the electronic device[.]” (*Id.* ¶¶[0066], [0153], [0053], [0075], [0077], Figs. 5A, 9; Ex. 1002 ¶42.) Shin’s device classified voice input based on distance and then tailored the output. For users within 1 meter, the device output “detailed content” at a lower volume (e.g., 40 dB). (Ex. 1003 ¶¶[0051], [0088], Tables 1, 3.) For users more than 1 meter away, it output “abbreviated content” at a higher volume (e.g., 45 dB or more). (*Id.*; Ex. 1002 ¶43.)

For example, if a user said, “Let me know today’s weather,” Shin’s device obtained weather-related information and may respond with: “The weather in Jul. 1,

2015 is (①) rainy after cloudy (②). The highest temperature is 28° C., and the lowest temperature is 18° C. (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).” (Ex. 1003 ¶¶[0088].) If the voice was classified as near (e.g., less than 1 meter away), all five categories of information were provided. (*Id.*) If the voice was classified as farther away (e.g., between 1 and 2 meters), category ⑤ was omitted. (*Id.*) And, if the voice was classified as even farther away (e.g., greater than 4 meters), all but category ② was omitted. (*Id.*) Thus, Shin disclosed using a proximity sensor to calculate user-to-device distance, classifying the user’s voice based on that distance, and tailoring the response’s content and volume based on the classification. This allowed Shin’s device to provide a “suitable” amount of information at a “suitable” volume based on distance. (*Id.* ¶¶[0089], [0080]; Ex. 1002 ¶44.)

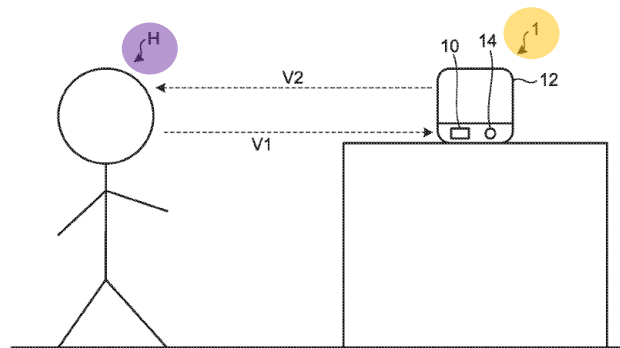
Other references also disclosed tailoring output responses and/or volume based on the user-to-device distance. (Ex. 1002 ¶¶36-40; Exs. 1011, 1012, 1013.)

III. THE ’337 PATENT

A. Overview

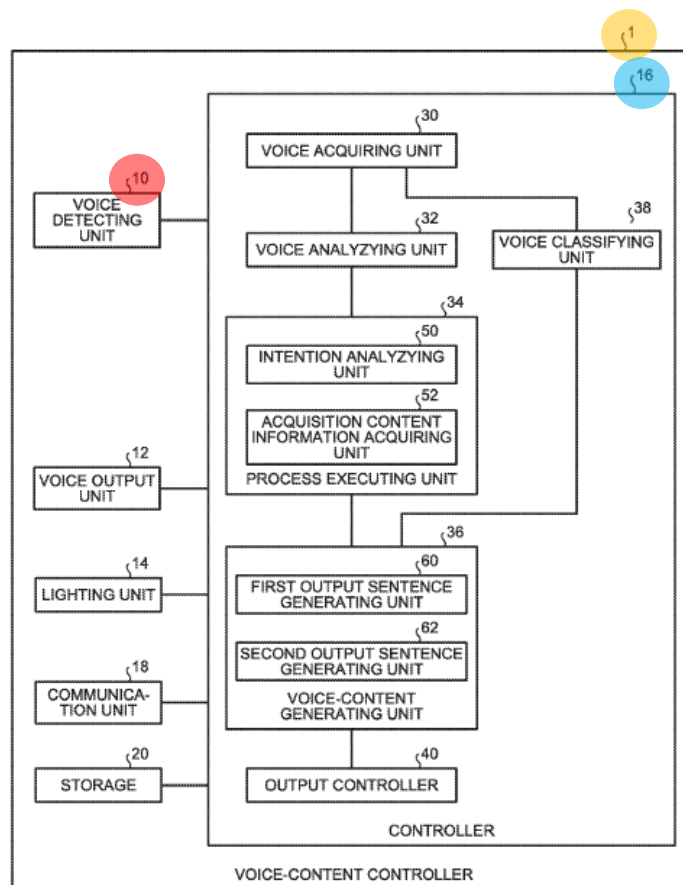
The ’337 patent describes a “voice-content control device” that analyzes a user’s voice and generates a response. (Ex. 1001, Abstract.) Figure 1 shows such a device 1 (orange) that detects the voice V1 of a user H (purple), processes it, and outputs a responsive voice V2:

FIG.1



(*Id.*, Fig. 1, 2:62-3:1; Ex. 1002 ¶45.)¹ The device 1 includes various components, such as a voice detecting unit 10 (red) and a controller 16 (blue):

FIG.2



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

(Ex. 1001, Fig. 2, 3:12-17; Ex. 1002 ¶46.) The controller includes several “units” for acquiring the voice, analyzing it, processing the request, and generating a response. (Ex. 1001, 3:50-59.) For example, the process executing unit analyzes the user’s speech (e.g., “How’s the weather today?”) detected with the voice detecting unit (e.g., microphones) and obtains the requested information (e.g., weather). (*Id.*, 4:12-14, 4:56-61, 7:17-27.)

The device may include a “proximity sensor” for calculating the distance to the user. (*Id.*, 8:20-26.) This generic sensor, described in a single sentence, purportedly allows the device to classify a voice as either a first or second voice. (*Id.*) When the voice is classified as a first voice (e.g., nearer than threshold distance), the device generates a first sentence and outputs it at a first volume. (*Id.*, 9:38-50, 13:62-65.) When the voice is classified as a second voice (e.g., farther than threshold distance), the device generates a second sentence that omits some information from the first sentence and outputs it at a different volume. (*Id.*; Ex. 1002 ¶47.)

B. Prosecution History

During prosecution, the applicant distinguished the prior art based on limitations that recite “a proximity sensor configured to calculate a distance between a user and the voice-content control device” and “a voice classifying unit configured ... to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device.” (Ex. 1007, 41-46.)

Apparently unaware that the prior art disclosed these limitations, the Examiner erroneously allowed the claims. (*Id.*, 31.)

IV. RELIEF REQUESTED

A. Grounds

The Board should cancel the claims on the following Grounds:

Ground	Reference(s)	Basis	Challenged Claims
1	Shin	§102	1-5
2	Shin	§103	1-5
3	Shimomura	§103	1-5
4	Shimomura and Shin	§103	1-5
5	Shin and/or Shimomura in view of Kristjansson	§103	1-5

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

B. The References Are Analogous Prior Art.

The '337 patent's earliest possible priority date is March 6, 2018. (Ex. 1001.)

Two references relied on herein are prior art under AIA §102(a)(1) and §102(a)(2):

- (1) Shin published March 23, 2017 (Ex. 1003); and
- (2) Shimomura published October 28, 2005 (Ex. 1004).

Kristjansson is prior art under AIA §102(a)(2) because it is an issued U.S. patent that was effectively filed no later than March 30, 2017 and names different inventors than the '337 patent (Ex. 1013).

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

Shin, Shimomura, and Kristjansson are analogous art because they are from the same field as the '337 patent, e.g., controlling a device's voice output. (Ex. 1001, 1:29-64, 3:7-11); *Unwired Planet, LLC v. Google Inc.*, 841 F.3d 995, 1000 (Fed. Cir. 2016). They are also pertinent to a problem the inventors were focused on, e.g., adjusting voice output to improve user interaction. (Ex. 1002 ¶25.)

V. LEVEL OF ORDINARY SKILL

Based on the relevant factors, *In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995), 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. (Ex. 1002 ¶¶30-34.) 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.

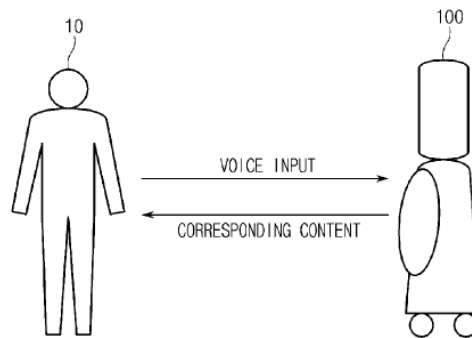
VII. GROUND 1: CLAIMS 1-5 ARE ANTICIPATED BY SHIN.

Shin anticipates claims 1-5. (Ex. 1002 ¶¶50-106.)

A. Claim 1

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

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



(*Id.*, Fig. 1A.) Thus, Shin discloses a “voice-content control device.” (Ex. 1002 ¶51.)

2. 1[a]: Proximity Sensor

Claim element 1[a] recites “a proximity sensor configured to calculate a distance between a user and the voice-content control device.” Shin discloses this. (Ex. 1002 ¶¶52-58.)

Shin’s device includes a “distance detection module” that “compute[s] a distance between a user and the electronic device[.]” (Ex. 1003 ¶[0066]; *see id.* ¶¶[0053] (“the processor 120 may determine a distance between the user and the electronic device 101 based on ... the distance computed, calculated, or measured by the distance detection module 180”), [0075], [0077], Fig. 5A; Ex. 1002 ¶53.) The distance detection module 180 may include “a depth camera like a time-of-flight (TOF) camera, a stereo camera computing depth information using triangulation, a charge coupled device (CCD) camera computing a distance through an image processing, or the like” and/or “various sensors,” such as “an infra-red sensor, an RF sensor, an ultrasonic sensor, and the like.” (Ex. 1003 ¶[0066]; *see also id.* ¶[0042];

Ex. 1002 ¶54.) Shin’s Figure 3 shows the electronic device 101 comprising the distance detection module 180 (green):

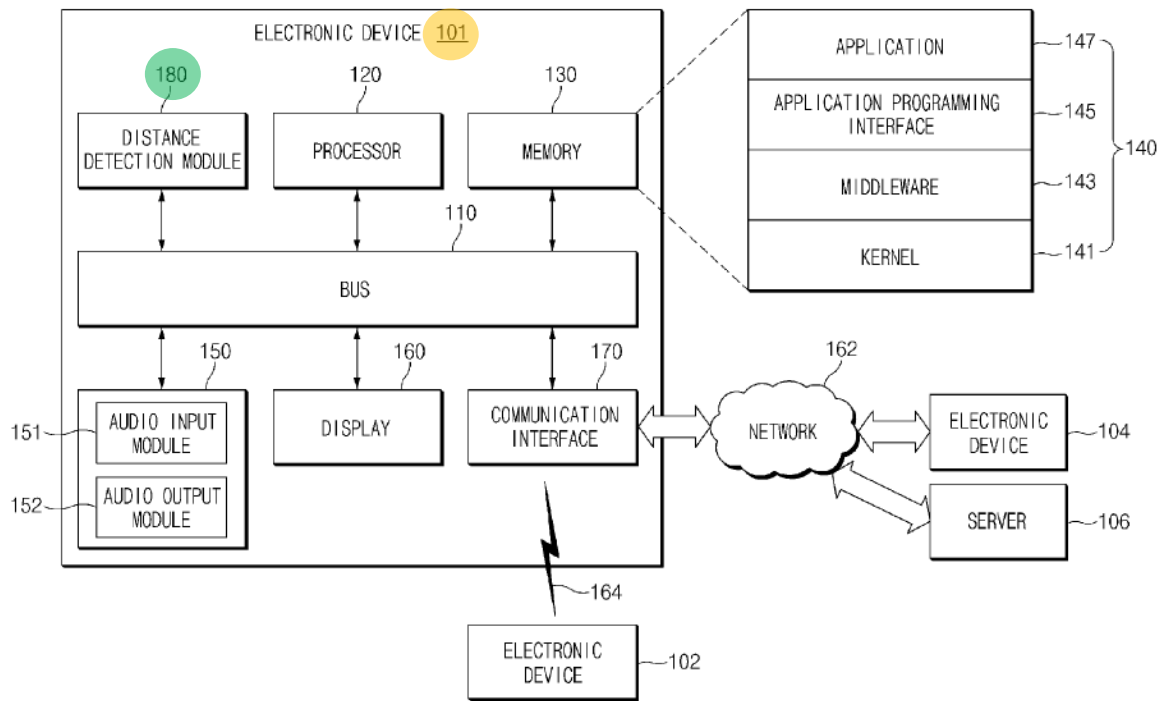


FIG. 3

(Ex. 1003, Fig. 3, ¶[0045]; *id.* ¶¶[0053]-[0054], [0077], [0096], claims 5, 9; Ex. 1002 ¶¶55-56.)²

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

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

Thus, Shin discloses a proximity sensor configured to calculate a distance between a user and the voice-content control device. (Ex. 1002 ¶¶52-58.)

3. 1[b]: Voice Classifying Unit

Claim element 1[b] recites “a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device.” Shin discloses this. (Ex. 1002 ¶¶59-66.)

The ’337 patent states that a voice classifying unit may be part of a controller, which performs processes by reading software/program stored in a storage. (Ex. 1001, 3:50-59, Fig. 2.) This controller can be a central processing unit (CPU). (*Id.*) Thus, according to the ’337 patent, the “voice classifying unit” refers to a CPU configured to perform the claimed function, namely, to analyze the user’s voice and to classify it as a first or second voice. (Ex. 1002 ¶60.)

The ’337 patent further explains that, when a proximity sensor is used for calculating the 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, 8:20-26.) Specifically, the 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.*, 8:30-34.) Shin discloses the same thing. (Ex. 1002 ¶61.)

First, Shin discloses a voice classifying unit that “analyze[s] a voice spoken by a user and acquired by a voice acquiring unit,” as recited. Specifically, Shin discloses that the device 101 (orange) includes a processor 120 (blue) and an audio input module 151 (red):

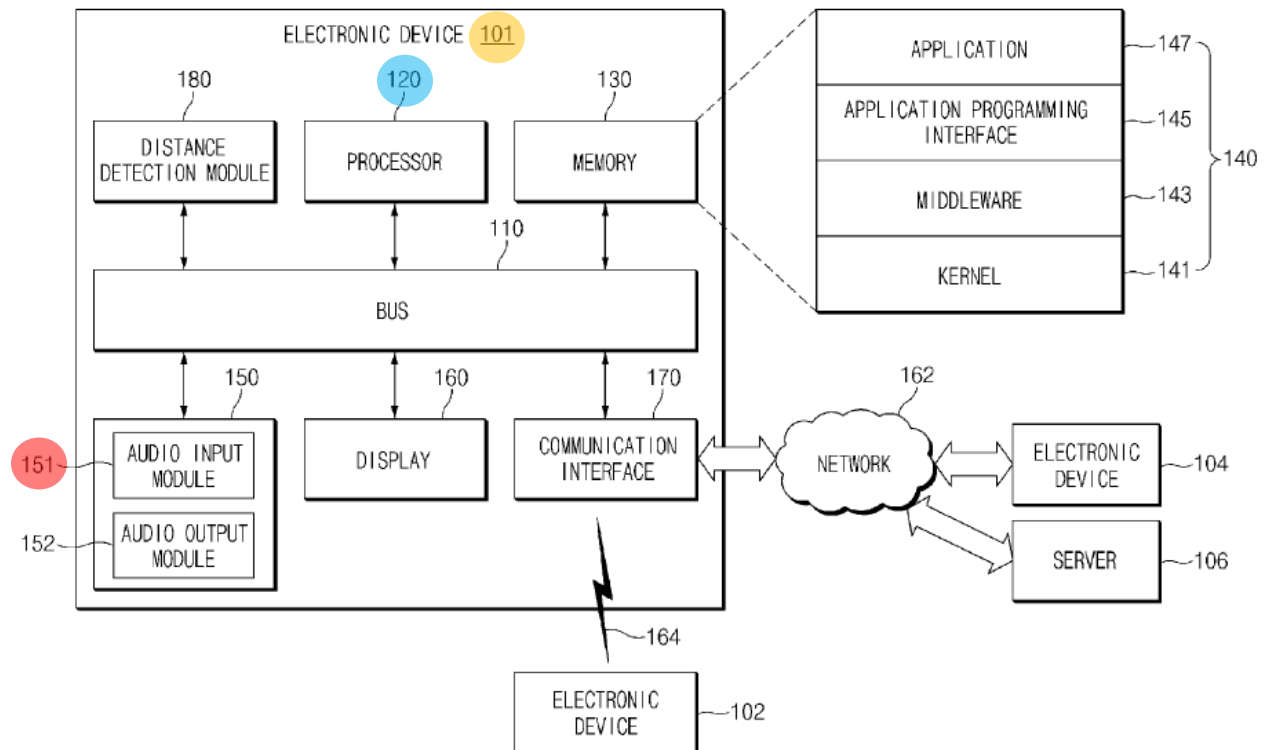


FIG. 3

(Ex. 1003, Fig. 3, ¶¶[0045]; Ex. 1002 ¶62.) The processor 120, which may be a CPU (Ex. 1003 ¶[0029]), “analyze[s] a voice input received through [the] audio input module 151.” (*Id.* ¶[0048]; *see also id.*, Abstract, ¶¶[0003]-[0011], [0037]-[0039], [0041], [0060]-[0061], claims 1, 15, 20.) The audio input module 151 can be “implemented with a microphone and the like” to “obtain a user’s speech as a voice input.” (*Id.* ¶[0062]; *see also id.* ¶¶[0037], [0155], [0158].) Thus, Shin discloses a

voice classifying unit (processor) configured to analyze a voice spoken by a user and acquired by a voice acquiring unit (e.g., audio input module and/or portion of processor that receives voice from the audio input module). (Ex. 1002 ¶62.)

Second, Shin discloses that the voice classifying unit (processor) analyzes the voice to classify it as either a first or second voice based on the user-to-device distance. (*Id.* ¶63.) Shin’s processor executes a voice recognition application to process the voice input and generates corresponding content for output. (Ex. 1003 ¶¶[0047], [0060]; *see also id.* ¶¶[0037], [0039].) The processor determines the “output scheme,” which includes the content and its volume level, “based on the distance between the user and the electronic device 101[.]” (*Id.* ¶¶[0078], [0086]; *see also id.* ¶¶[0053], [0077]-[0080], claims 6, 17.) To do so, Shin’s processor classifies the voice based on the distance. (Ex. 1002 ¶63.)

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

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

(Ex. 1003, Table 3 (consolidated).)

Thus, for example, Shin discloses classifying the voice as a “first voice” when the distance is less than 1 meter, and a “second voice” when the distance is greater than 4 meters:

TABLE 3		
	Distance (D) between user and electronic device	Output amount of information of content
First Voice	D < 1 m	Amount of information of original content
	1 m ≤ D < 2 m	Content of first abbreviation level
	2 m ≤ D < 3 m	Content of second abbreviation level
	3 m ≤ D < 4 m	Content of third abbreviation level
Second Voice	D ≥ 4 m	Content of fourth abbreviation level (abstract content)

(*Id.*; see also *id.*, Table 1, ¶¶[0051] (content may be “classified dichotomously”), [0079]-[0081], [0086]-[0087]; Ex. 1002 ¶64.) Any of the distance ranges up to 4 meters would satisfy the claimed “first voice.” (Ex. 1002 ¶65.) For example, a voice between 3 and 4 meters away could be considered the recited “first voice.” (*Id.*) Alternatively, all of the distances less than 4 meters could collectively be considered a first voice (e.g., a voice less than 4 meters away). (*Id.*)

Thus, Shin discloses that the voice classifying unit (processor) is configured to analyze the voice to classify the voice as either one of a first voice (e.g., closer than 1 meter) or a second voice (e.g., farther than 4 meters) based on the distance between the user and the device. (*Id.* ¶¶59-66.)

4. 1[c]: Process Executing Unit

Claim element 1[c] recites “a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user.” The ’337 patent explains that, like the voice classifying unit, a process executing unit may be part of a controller that performs processes by reading software/program stored in a storage. (Ex. 1001, 3:50-59, Fig. 2.) In the ’337 patent, the “process executing unit” analyzes the user’s speech (e.g., “How’s the weather today?”) and obtains the requested information. (*Id.*, 4:12-14, 4:56-61, 7:17-27.) The patent admits that devices that performed these functions were known (*id.*, 1:21-28; Ex. 1002 ¶¶67) and, consequently, this limitation cannot make the claim patentable. Regardless, Shin discloses this claim element. (Ex. 1002 ¶¶67-70.)

Shin’s processor analyzes the voice acquired by the voice acquiring unit (e.g., the “audio input module” and/or portion of processor that receives voice from the audio input module). (Ex. 1003 ¶¶[0048], [0062].) Shin’s processor executes a voice recognition application to process the voice input (which requests content in which the user is interested) and generates corresponding content for output by analyzing the request. (*Supra* §VII.A.3; Ex. 1003 ¶¶[0048], [0058]-[0060], [0084], [0089]; Ex. 1002 ¶68; *see also* Ex. 1003 ¶¶[0037], [0058], [0067], [0071], [0088], [0103], [0107], [0120], [0126]-[0127], [0141].)

Indeed, Shin discloses analyzing the acquired voice “to execute processing required by the user” in the same way as the ’337 patent. (Ex. 1002 ¶¶69.) For example, Shin describes analyzing a user’s speech (“Let me know today’s weather”) to execute processing to obtain the required weather information. (*Id.*; Ex. 1003 ¶[0088].)

Thus, Shin discloses a process executing unit (processor) configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user. (Ex. 1002 ¶¶67-70.)

5. 1[d]: Voice-Content Generating Unit

Claim element 1[d] recites “a voice-content generating unit configured to generate, based on content of the processing executed by the process executing unit, output sentence that is text data for a voice to be output to the user.” In the ’337 patent, the “voice-content generating unit” is the portion of the processor (CPU) that performs the recited function. (Ex. 1001, 3:50-59, Fig. 2.) In the ’337 patent’s weather example, the “voice-content generating unit” performs this function when it generates the following output sentence: “On the day Z of the month Y of the year X, in Tokyo, the weather is partly cloudy, the chance of rainfall is 20%, the highest air temperature is 25 degrees, and the lowest air temperature is 15 degrees.” (*Id.*, 9:16-21.) Again, the patent admits that devices that performed these functions were

known (*id.*, 1:21-28; Ex. 1002 ¶71) and, consequently, this limitation cannot make the claim patentable. Regardless, Shin discloses this element. (Ex. 1002 ¶¶71-73.)

Shin discloses a voice-content generating unit that generates an output sentence as text data for a voice to be output to the user. (*Id.* ¶72.) As discussed above, Shin’s processor generates the content for output to the user. (*Supra* §§VII.A.3-VII.A.4.) For example, Shin discloses that the content can be a sentence including text data, such as: “The current time is nine ten AM.” (Ex. 1003 ¶[0037].) In another example, Shin’s processor performs the claimed function when it generates the output: “The weather in Jul. 1, 2015 is (①) rainy after cloudy (②). The highest temperature is 28° C., and the lowest temperature is 18° C. (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).” (*Id.* ¶[0088].) For example, the output sentence may be “The weather in Jul. 1, 2015 is rainy after cloudy.” (*Id.*; Ex. 1002 ¶72.) This content may be converted into a voice using text-to-speech technology and output audibly. (Ex. 1003 ¶[0048]; *see also id.* ¶¶[0007], [0037], [0039], [0062], [0072], [0109], [0131], [0139], [0158].) Thus, Shin discloses a voice-content generating unit (processor) configured to generate an output sentence that is text data for a voice to be output to the user. (Ex. 1002 ¶¶71-72.)

Shin also discloses the voice-content generating unit generates the output sentence based on content of the processing executed by the process executing unit. As discussed above, Shin’s processor analyzes the voice input to generate content which

aligns with the user’s interests. (*Supra* §VII.A.4.) Thus, Shin discloses a voice-content generating unit (processor) configured to generate, based on content (e.g., weather information) of the processing executed by the process executing unit (processor), output sentence that is text data for a voice to be output to the user. (Ex. 1002 ¶73.)

6. 1[e]: Output Controller

Claim element 1[e] recites “an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence therein.” Shin discloses this claim element. (Ex. 1002 ¶¶74-75.)

Shin’s processor converts the content to voice (*supra* §VII.A.5), and can “adjust” the “output volume level” (Ex. 1003 ¶[0053]; *see also id.* ¶¶[0076]-[0081], [0098]-[0101], [0105], claims 6, 17). Thus, Shin discloses an output controller (processor) configured to adjust a sound volume of voice data obtained by converting the output sentence therein. (Ex. 1002 ¶75; *infra* §VII.A.9.)

7. 1[f]: Generate a First Output Sentence

Claim element 1[f] recites that “the voice-content generating unit is further configured to generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice[.]” Shin discloses this claim element. (Ex. 1002 ¶¶76-79.)

Shin’s processor adjusts the amount of information provided to the user based on the user-to-device distance. (Ex. 1003 ¶[0053]; *see also id.* ¶¶[0051], [0086]-[0090], claims 8, 17; Ex. 1002 ¶77.) Specifically, when a voice is classified as a first voice (e.g., closer than 1 meter), the device generates a first 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 corresponding to a voice input”), [0051].) The amount of information provided is shown in Table 3:

TABLE 3	
Distance (D) between user and electronic device	Output amount of information of content
First Voice — D < 1 m	Amount of information of original content
1 m ≤ D < 2 m	Content of first abbreviation level
2 m ≤ D < 3 m	Content of second abbreviation level
3 m ≤ D < 4 m	Content of third abbreviation level
D ≥ 4 m	Content of fourth abbreviation level (abstract content)

(*Id.*, Table 3; Ex. 1002 ¶77.) Thus, when classified as a first voice (e.g., closer than 1 meter), the processor outputs the content without abbreviation. (Ex. 1003 ¶[0087]; Ex. 1002 ¶77.)

Shin discloses an example in which the “original content,” and thus the output when the voice is classified as the first voice (e.g., closer than 1 meter), is: “The weather in Jul. 1, 2015 is (①) rainy after cloudy (②). The highest temperature is

28° C., and the lowest temperature is 18° C. (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).” (Ex. 1003 ¶[0088].) In this example, the first output sentence may be “The weather in Jul. 1, 2015 is rainy after cloudy.” (*Id.*; Ex. 1002 ¶78.)

Thus, Shin discloses that the voice-content generating unit (processor) is further configured to generate a first output sentence (e.g., containing information of original content) as the output sentence when the acquired voice has been classified as the first voice (e.g., closer than 1 meter). (Ex. 1002 ¶¶76-79.)

8. 1[g]: Generate a Second Output Sentence

Claim element 1[g] recites that the voice-content generating unit is further configured to “generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice.” Shin discloses this claim element. (Ex. 1002 ¶¶80-85.)

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.” (Ex. 1003 ¶[0087]; *id.* ¶[0086] (“output amount of information of the content decreases as a distance between the user and the electronic device 101 increases”).) The varying abbreviation levels are shown in Table 3:

TABLE 3

	Distance (D) between user and electronic device	Output amount of information of content
First Voice	D < 1 m	Amount of information of original content
	1 m ≤ D < 2 m	Content of first abbreviation level
	2 m ≤ D < 3 m	Content of second abbreviation level
	3 m ≤ D < 4 m	Content of third abbreviation level
Second Voice	D ≥ 4 m	Content of fourth abbreviation level (abstract content)

(*Id.*, Table 3; Ex. 1002 ¶81.) As shown, when the user is more than 4 meters from the device, the voice falls into a second classification and the processor outputs “[c]ontent of fourth abbreviation level (abstract content).” (Ex. 1003 ¶[0087].) Thus, when the voice is classified as a second voice (e.g., greater than 4 meters), the output content omits some information provided when the voice is classified as a first voice. (*Id.* ¶[0051] (content may be classified into a second category that includes the “least” amount of information); *id.* (the processor may “extract[]” a portion of the content); Ex. 1002 ¶81.)

In the weather example, Shin’s device provides information ①-⑤ when the voice is classified as a first voice. (Ex. 1003 ¶[0088].) That includes generating the first output sentence “The weather in Jul. 1, 2015 is rainy after cloudy.” (*Id.*; Ex. 1002 ¶82.) In contrast, a voice from more than 4 meters away is classified as a second voice and the processor generates the second output sentence “rainy after cloudy.” (Ex. 1003 ¶[0088] (“the content of a fourth abbreviation level (abstract

content) may include only information of (2)”; Ex. 1002 ¶82.) Thus, Shin discloses omitting some information (e.g., (1), “The weather in Jul. 1, 2015 is”) when the voice is classified as the second voice. (*Id.*; *see also* Ex. 1002 ¶¶83-84.)

Thus, Shin discloses that the voice-content generating unit (processor) is further configured to generate a second output sentence (e.g., content of a fourth abbreviation level) in which information is omitted (e.g., information of (1)) as compared to the first output sentence (e.g., information of original content) as the output sentence when the acquired voice has been classified as the second voice (e.g., greater than 4 meters away). (Ex. 1002 ¶¶80-85.)

9. 1[h]: Output Controller Adjusts Volume of Voice Data

Claim element 1[h] recites that “the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto.” Shin discloses this claim element. (Ex. 1002 ¶¶86-90.)

Shin’s processor adjusts the content’s output volume based on the user-to-device distance. (*Supra* §VII.A.6; Ex. 1003 ¶¶[0053], [0077]-[0081].) Shin discloses that the “device 101 may set an output volume level of the content to 40 dB when the distance between the user and the electronic device 101 is less than 1 m

and may increase the output volume level by 5 dB whenever the distance increases by 1m.” (Ex. 1003 ¶[0080].) Shin’s Table 1 also discloses this:

TABLE 1

Distance (D) between user and electronic device	Output volume level of content
$D < 1 \text{ m}$	40 dB
$1 \text{ m} \leq D < 2 \text{ m}$	45 dB
...	...
$D \geq 5 \text{ m}$	65 dB

(*Id.*, Table 1.) Thus, Shin discloses that the output volume when the distance is between 4 and 5 meters may be 60 dB. (*Id.* ¶[0080], Table 1; Ex. 1002 ¶87.)

Considering Shin’s disclosure regarding the output information and volume together, Shin discloses that when a voice is classified as a first voice (e.g., closer than 1 meter), a first output sentence containing the “original content” is provided (Table 3) at a volume level of 40 dB (Table 1). (Ex. 1002 ¶88.) When a voice is classified as a second voice (e.g., greater than 4 meters), a second output sentence containing “content of fourth abbreviation level” is provided (Table 3) at a volume level of 60 or 65 dB (Table 1 and ¶[0080]). (*Id.*) Shin repeatedly discloses that both content and volume may vary based on distance. (Ex. 1003 ¶¶[0049], [0053], [0140]; Ex. 1002 ¶88.)

Thus, Shin discloses that the output controller (processor) is further configured to adjust the sound volume of voice data such that the sound volume (e.g., 40 dB) of voice data obtained by converting the first output sentence (e.g., information

of original content) thereinto differs from the sound volume (e.g., 60 or 65 dB) of the voice data obtained by converting the second output sentence (e.g., content at fourth abbreviation level) thereinto. (Ex. 1002 ¶¶89.)

Accordingly, Shin anticipates claim 1. (*Id.* ¶¶51-90.)

B. Claim 2

Shin discloses claim 2’s two additional claim elements. (Ex. 1002 ¶¶91-99.) Moreover, these elements reflect the operation of known voice processing systems, and therefore cannot render the claim patentable. (*Id.* ¶91.)

1. 2[a]: Process Executing Unit

Claim element 2[a] recites that the “process executing unit” comprises: [i] “an intention analyzing unit configured to extract intention information indicating an intention of the user based on the voice acquired by the voice acquiring unit; and [ii] an acquisition content information acquiring unit configured to acquire acquisition content information which is notified to the user based on the extracted intention information[.]” The ’337 patent discloses that the “intention analyzing unit” extracts intention information indicating the user’s intent when, for example, it analyzes the phrase “How’s the weather today?” and recognizes that the user wants weather-related information. (Ex. 1001, 4:56-66.) The “information acquiring unit” in the ’337 patent acquires content information that is notified to the user based on the extracted intent information when, for example, it acquires the weather-related information.

(*Id.*, 7:14-27, 7:28-36, 9:9-16 (intention information is “weather” and acquisition information is “partly cloudy”); Ex. 1002 ¶92.)

As discussed, Shin’s processor analyzes the voice to identify the content the user is interested in, and then generates the requested content. (*Supra* §VII.A.4.) Shin expressly discloses that the device can “receive the voice input ... and may generate content corresponding to a result of analyzing the voice input.” (Ex. 1003 ¶[0037].) For example, in response to a user’s statement, “Let me know what time it is now,” the device “may generate content, such as ‘The current time is nine ten AM[.]’” (*Id.*; *see also id.* ¶[0039].) Shin further discloses that, in response to the voice input, “Let me know today’s weather,” the device generates a response containing weather-related information, such as “cloudy.” (*Id.* ¶[0088].) To obtain such content, the device may include an application that “analyz[es] a voice input[] and generat[es] and/or provid[es] the content corresponding to the voice input.” (*Id.* ¶[0060].) The device can also delegate some processing to another device (e.g., another electronic device or a server), which then delivers the results. (*Id.* ¶[0067]; *see also id.* ¶¶[0064]-[0065], [0174]-[0176].) Accordingly, Shin discloses acquiring the information from memory or from another device. (Ex. 1002 ¶93.)

Even if Shin had not expressly disclosed this claim element, Shin inherently discloses it. Shin’s processor necessarily includes an “intention analyzing unit” and an “acquisition content information acquiring unit” because such units would have

been required for Shin’s device to function as described. (*Id.* ¶94.) 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 information (e.g., “cloudy” or “nine ten AM”). (*Id.*) These functionalities are required for the processor to obtain and deliver the requested content as Shin describes. (*Id.*)

Thus, Shin expressly or inherently discloses that the process executing unit (processor) comprises: an intention analyzing unit configured to extract intention information (e.g., weather or time) indicating an intention of the user based on the voice acquired by the voice acquiring unit. (*Id.* ¶95.) Shin also expressly or inherently discloses that the process executing unit (processor) comprises an acquisition content information acquiring unit configured to acquire acquisition content information (e.g., weather-related information or time information) which is notified to the user based on the extracted intention information (e.g., weather or time). (*Id.*)

Moreover, even if Shin had not disclosed this claim element, it would not make the claim patentable because the ’337 patent admits this claim element was known. (Ex. 1001, 1:21-28 (known devices analyze user voice, perform processing “according to the user’s intention,” and output, via voice, content of a user’s inquiry); Ex. 1002 ¶96.)

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

Claim element 2[b] recites that “the voice-content generating unit is further configured to generate the text data including the acquisition content information as the output sentence.”

Shin’s processor acquires information, such as time-related or weather-related information, for generating a response to the user. (*Supra* §VII.B.1; Ex. 1003 ¶¶[0037], [0039], [0088].) This information is incorporated into the content provided to the user, as demonstrated by the examples discussed for claim element 2[a]. (*Id.*; Ex. 1002 ¶98.) The content is provided to the user “as sound based on a text to speech (TTS) technology.” (Ex. 1003 ¶[0048].) Thus, Shin discloses that the voice-content generating unit (processor) is configured to generate the text data including the acquisition content information (e.g., weather-related information or time information) as the output sentence. (Ex. 1002 ¶98.)

Accordingly, Shin anticipates claim 2. (*Id.* ¶¶91-99.)

C. Claim 3

Claim 3 depends from claim 1 and further recites “wherein, on generating the second [output] sentence, the voice-content generating unit is further configured to omit a part of information included in the voice spoken by the user.” Shin discloses this. (*Id.* ¶¶100-03.)

For example, Shin discloses classifying a voice received from a distance of 4 meters or more as a second voice. (*Supra* §VII.A.8; Ex. 1003, Table 3.) For that second voice, in response to the request “Let me know today’s weather,” the second output sentence is: “rainy after cloudy.” (*Id.*; Ex. 1003 ¶[0088].) This output omits all user-spoken words, as well as information such as “Jul. 1, 2015” (corresponding to “today’s”) included in the user’s voice. (*Id.*; *see also* Ex. 1003 ¶[0039] (the input “What time is it now?” results in output “nine ten,” omitting all words and information included in the voice spoken by the user); Ex. 1002 ¶101.) Thus, Shin discloses, on generating the second output sentence, the voice-content generating unit (processor) is further configured to omit a part of information included in the voice spoken by the user. (Ex. 1002 ¶101.)

Alternatively, Shin’s processor may generate the second output sentence to include content at a first level of abbreviation when the acquired voice is received from a distance between 1 and 2 meters away. (Ex. 1003, Table 3.) This first level of abbreviation also omits part of the information included in the voice spoken by the user. (Ex. 1002 ¶102.) For example, in response to the request, “Let me know today’s weather” (Ex. 1003 ¶[0088]), the processor outputs information that omits the information “let me know.” (*Id.*; Ex. 1002 ¶102.) Accordingly, Shin anticipates claim 3. (Ex. 1002 ¶¶100-03.)

D. Claims 4 and 5

The preamble of claim 4 recites “[a] voice-content control method.” Shin discloses both electronic devices and methods of operating the electronic devices. (Ex. 1003 ¶[0181]; Ex. 1002 ¶104; *see also* Ex. 1003 ¶¶[0003], [0008]-[0011], [0016]-[0021], [0043], [0047], [0092]-[0093], Title, Abstract, claims 15-19.)

The preamble of claim 5 recites “[a] non-transitory storage medium that stores a voice-content control program that causes a computer to execute” a series of steps, which are the same as the steps recited in claim 4. Shin discloses a non-transitory storage medium. Specifically, Shin discloses electronic devices implemented by instructions stored in a computer-readable storage media. (Ex. 1003 ¶[0181]; Ex. 1002 ¶105; *see also id.* ¶¶[0011], [0179]-[0180], [0182], claim 20.)

After the preamble, the remainder of claims 4 and 5 are identical and merely recite the functions performed by the various units recited in claim 1. For example, claims 4 and 5 recite “calculating a distance” (performed in claim 1 by the proximity sensor, *supra* §VII.A.2), “acquiring a voice” (claim 1’s voice acquiring unit, *supra* §VII.A.3), “analyzing the acquired voice” (claim 1’s voice classifying unit and process executing unit, *supra* §§VII.A.3-VII.A.4), “generating ... [an] output sentence” (claim 1’s voice-content generating unit, *supra* §§VII.A.5, VII.A.7-VII.A.8), and “adjusting a sound volume” (claim 1’s output controller, *supra* §§VII.A.6, VII.A.9). As discussed above, Shin discloses that the electronic device performs each of these

functions and therefore Shin anticipates claims 4 and 5. (*Supra* §VII.A; Ex. 1002 ¶106.)

VIII. GROUND 2: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIN.

Shin anticipates claims 1-5 for the reasons set forth above. (*Supra* §VII.) If PO argues that Shin fails to disclose any limitations, such limitations would have been obvious over Shin. (Ex. 1002 ¶¶107-28.)

A. Claim 1

Even if PO were to argue that claim 1 requires the user’s voice to be classified into no more than two classifications based on distance, Shin renders this obvious. (Ex. 1002 ¶¶108-16.) A POSITA would have been motivated for several reasons to implement Shin’s volume change table (Table 1) and amount of information table (Table 3) with only two distance ranges. (*Id.*)

First, Shin suggests doing so. Shin discloses that content can be divided into two classifications. (Ex. 1003 ¶[0051].) Based on that disclosure, a POSITA would have been motivated to also use two volume levels, one for each classification. (Ex. 1002 ¶109.)

Second, Shin discloses that its “embodiments ... are not limited to the examples of Table 1, Table 2, and Table 3.” (Ex. 1003 ¶[0090]; *see also id.* ¶[0081].) 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 to adjust content output based on user distance. (Ex. 1002 ¶110.)

Third, simplifying complex systems into fewer categories is a routine and well-established approach to reduce computational load and improve efficiency. (*Id.* ¶111); *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007). Binary classification (classifying into 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 ¶111.) 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 described needs. (*Id.*)

Fourth, such a modification reflects merely applying a known technique (binary classification) to achieve predictable results (a streamlined, efficient, and straightforward user interface with content output adjusted based on “near” or “far” proximity classifications). (*Id.* ¶112); *KSR*, 550 U.S. at 417.

Fifth, such a modification reflects merely applying a known technique (binary classification) to a known device (Shin’s) that is ready for improvement and yields predictable results (consistent and reliable adjustments to content detail and volume based on user proximity, enhancing user interaction). (Ex. 1002 ¶113); *KSR*, 550 U.S. at 417.

Sixth, such a modification would have been obvious to try. (Ex. 1002 ¶114); *KSR*, 550 U.S. at 417. Shin’s system accommodates adaptable classifications, and a POSITA would have reasonably expected success in using binary classification to meet Shin’s functional goals while simplifying implementation and minimizing computational demand. (Ex. 1002 ¶114.)

A POSITA would have reasonably expected success in modifying Shin in this way. (*Id.* ¶115.) Doing so would have been trivial, involving configuring Shin’s processor with revised tables for two proximity-based categories and optimizing performance for specific uses. (*Id.*)

Accordingly, Shin renders obvious claim 1. (*Id.* ¶¶108-16.)

B. Claim 2

Even if Shin had not disclosed claim 2’s additional limitations, claim 2 would have been obvious over Shin. (Ex. 1002 ¶¶117-20.) Shin’s processor analyzes voice input to identify content the user is interested in, and then acquires and provides that content. (*Supra* §§VII.A.4, VII.B.1.) A POSITA would have recognized that implementing an “intention analyzing unit” to extract the user’s intent, alongside an “acquisition content information acquiring unit” to retrieve content based on the extracted intention, would have been a predictable and logical extension of Shin’s functionalities. (Ex. 1002 ¶117.)

Shin describes voice processing capabilities enabling the device to interpret user intent, such as identifying requests for time or weather, and generate appropriate responses. (*Supra* §VII.B.1.) A POSITA would have been motivated to structure the processor with an intention analyzing unit and a content acquisition unit for greater efficiency. (Ex. 1002 ¶118.) This modular design would facilitate consistent intent extraction, reliable content retrieval, and improved clarity and organization of the device’s processes. (*Id.*) It would merely reflect the way voice assistants worked at the time of the invention. (*Id.* (citing Ex. 1015).) Thus, it merely reflects known voice analysis techniques. (*Id.*; *see also infra* §XI.)

This approach would have been straightforward for a POSITA, as Shin’s device already integrates voice analysis and content generation. (*Id.* ¶119.) Implementing these functionalities as processor units would have been an obvious design choice, enhancing modularity, maintainability, and troubleshooting. (*Id.*) Using these units, Shin’s device could improve response accuracy and handle diverse user requests with greater flexibility. (*Id.*)

Accordingly, Shin renders obvious claim 2. (*Id.* ¶¶117-20.)

C. Claim 3

Even if Shin had not disclosed claim 3’s additional limitations, this claim would have been obvious over Shin. (Ex. 1002 ¶¶121-27.) A POSITA would have

been motivated for several reasons to configure Shin’s processor to selectively omit parts of a user’s spoken information in an output sentence. (*Id.*)

First, this would have been a predictable and desirable implementation of Shin’s processor, which processes and responds to user inputs. (*Id.* ¶122); *KSR*, 550 U.S. at 417. A POSITA would have understood that repeating all user-spoken information (e.g., the user speaks, “What time is it?” and the response begins with “The time is ...”) would be unnecessary. (Ex. 1002 ¶122.) Avoiding repetition improves efficiency and prevents response delays. (*Id.*)

Second, market demands and user expectations for efficient, streamlined interactions would have motivated this implementation. (*Id.* ¶123); *KSR*, 550 U.S. at 417. Users expect minimal delays and concise responses. (Ex. 1002 ¶123.) Omitting user-spoken information would align Shin’s processor with these expectations, improving user satisfaction. (*Id.*) A POSITA would have recognized this as an obvious solution. (*Id.*)

Third, such an implementation would have been a matter of routine experimentation and common sense. (*Id.* ¶124); *KSR*, 550 U.S. at 417. Given Shin’s aims to enhance user interactions (Ex. 1003 ¶¶[0007]-[0009], [0037], [0140]-[0141]), reducing unnecessary repetition would have been a logical step and would predictably improve performance without requiring significant innovation. (Ex. 1002 ¶124.)

Fourth, such an implementation would have been obvious to try. (*Id.* ¶125); *KSR*, 550 U.S. at 417. Shin’s device presents only two design options for responding: include all information from a user’s request or less than all. (Ex. 1002 ¶125.) Omitting some information to enhance efficiency and conciseness would have been an obvious choice. (*Id.*)

A POSITA would have reasonably expected success in implementing Shin’s processor to omit some user-spoken information when responding. (*Id.* ¶126.) This trivial implementation would have involved straightforward changes to configure the processor for more efficient and streamlined interactions. (*Id.*)

Accordingly, Shin renders obvious claim 3. (*Id.* ¶¶121-27.)

D. Claims 4 and 5

Claims 4 and 5 would have been obvious over Shin for the reasons explained regarding claim 1 in this Ground. (Ex. 1002 ¶128; *supra* §§VII.D, VIII.A.)

IX. GROUND 3: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA.

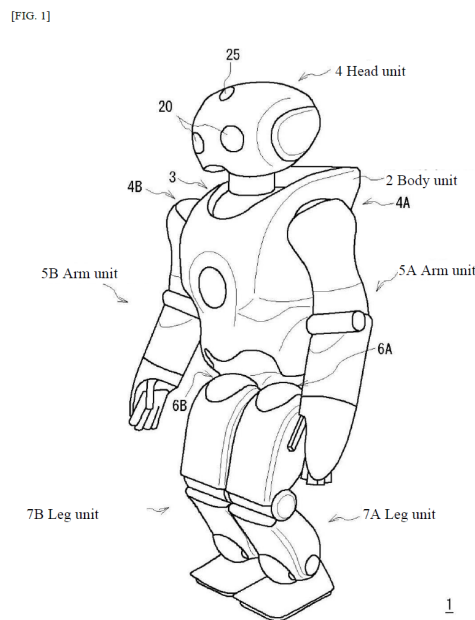
Shimomura renders obvious claims 1-5. (Ex. 1002 ¶¶129-97.)

A. Claim 1

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

Shimomura discloses a “speech control device that controls the speech of a device[.]” (Ex. 1004, claim 1; *see also id.*, Title.) Shimomura discloses an electronic

device, such as a robot or other interactive device, with input and output units corresponding to human sensory systems. (*Id.* ¶¶[0020]-[0022], [0096].) Inputs include “a pair of CCD (Charge Coupled Device) cameras 20” (eyes) and “a pair of microphones 21” (ears). (*Id.* ¶[0021].) Outputs include “a speaker 23” (mouth). (*Id.* ¶[0022].) Figure 1 shows an example:



(*Id.*, Fig. 1; Ex. 1002 ¶130.)

The device receives speech via microphones, performs speech recognition, and determines the “content to be spoken[.]” (Ex. 1004 ¶[0046]; *see also id.*, Abstract, Title, ¶¶[0001], [0030], Fig. 4.) For example, if a user asks about the weather, the device may respond, “Today’s weather in Tokyo is sunny.” (*Id.* ¶[0065]; Ex. 1002 ¶131.) The device “controls speech forms such as speech volume ... according

to the physical distance to the conversation partner.” (Ex. 1004 ¶¶[0049]-[0051]; *see also id.* ¶¶[0009]-[0010], [0016], [0057], [0085], [0088], [0090], [0092], claims 1-24; Ex. 1002 ¶132.)

Thus, Shimomura discloses a “voice-content control device.” (Ex. 1002 ¶¶130-32.)

2. 1[a]: Proximity Sensor³

Shimomura discloses that its device includes a “distance detection means for detecting the distance to the conversation partner user[.]” (Ex. 1004 ¶[0093].) This may include “a pair of CCD cameras 20 ... that detects the distance by a stereo vision method” or a “distance sensor[.]” (*Id.*; *id.* ¶¶[0039], [0059], [0070], claims 1 (“a distance detection means for detecting the distance between the device and the user”), 9, 17.)

Thus, Shimomura discloses a proximity sensor (image recognition unit, distance detection means, or distance sensor) configured to calculate a distance between a user and the voice-content control device. (Ex. 1002 ¶¶133-35.)

3. 1[b]: Voice Classifying Unit

First, Shimomura discloses a voice classifying unit that is configured to “analyze a voice spoken by a user and acquired by a voice acquiring unit” as recited.

³ Because the claim language is recited in Ground 1, it is not repeated here.

Shimomura discloses that its device, robot 1 (orange), includes a control unit 10 (blue) and a microphone 21 (red):

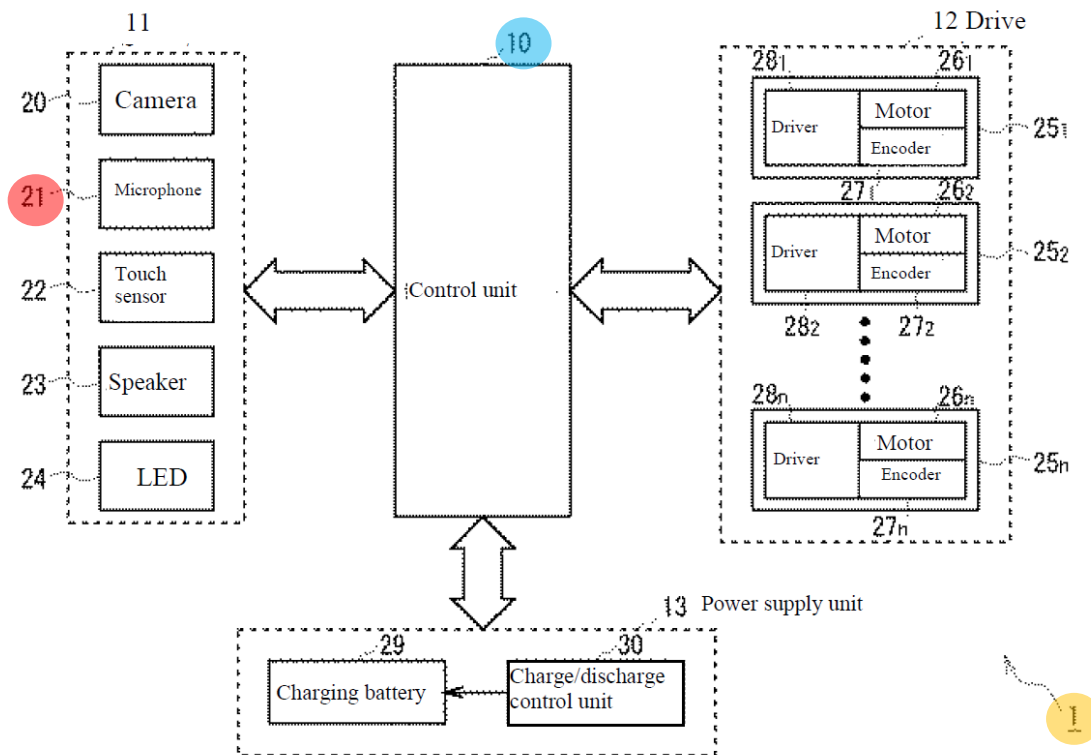


FIG. 2 Hardware constitution of the robot according to this embodiment (2)

(Ex. 1004, Fig. 2; Ex. 1002 ¶137.) The control unit 10, which includes a CPU, controls overall operation and other data processing. (Ex. 1004 ¶¶[0020], [0027], Fig. 3; *see also id.* ¶¶[0028], [0032], [0034].) The control unit implements a “behavior control system” that includes a “speech recognition unit.” (*Id.* ¶¶[0036]-[0038], [0042], Fig. 4; Ex. 1002 ¶137.) “The speech recognition unit 42 performs various sound-related recognition processes such as speech recognition, speaker recognition, and sound source direction recognition based on the speech signal S2 provided by

the microphone 21.” (Ex. 1004 ¶[0040].) Thus, Shimomura discloses a voice classifying unit (control unit) that is configured to analyze a voice spoken by a user and acquired by a voice acquiring unit (speech recognition unit and/or portion of processor that receives voice from microphone). (Ex. 1002 ¶137.)

Second, Shimomura discloses that the voice classifying unit (control unit) analyzes the voice to classify it as either a first or second voice based on the user-to-device distance. (Ex. 1002 ¶138.) Shimomura’s control unit also includes a “speech synthesis unit” and a “speech character string transformation unit.” (Ex. 1004 ¶¶[0036]-[0037], [0046], [0058], [0067], Fig. 4; Ex. 1002 ¶138.) These units may “transform the character string” of the content spoken by the robot or change its volume “according to the distance to the conversation partner.” (Ex. 1004 ¶¶[0058], [0067]-[0070]; *see also id.* ¶¶[0060]-[0063], [0073]-[0084], Abstract, claims 1-24.) To do so, Shimomura’s control unit first classifies the voice. (Ex. 1002 ¶138.)

For example, Shimomura discloses classifying voice 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. (Ex. 1002 ¶139.) 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.*, Fig. 7, ¶¶[0060]-[0064]; Ex. 1002 ¶139.)

Thus, Shimomura discloses that the voice classifying unit (processor) is configured to analyze the voice to classify the voice as either one of a first voice (e.g., farther than 350 cm) or a second voice (e.g., closer than 350 cm) based on the distance between the user and the device. (Ex. 1002 ¶¶136-40.)

4. 1[c]: Process Executing Unit

Shimomura's control unit is configured to analyze the voice acquired by the voice acquiring unit. (Ex. 1004 ¶[0040]; Ex. 1002 ¶142.) It recognizes speech from the microphone's signal and, based on that speech, determines the content to be spoken to the user. (*Supra* §IX.A.3.) Shimomura describes the user as a conversation partner of the device (Ex. 1004 ¶¶[0050], [0054], [0057]-[0059], [0067], [0086]-[0087], [0090], claims 1, 9, 17; *see also id.* ¶¶[0003]-[0004]), confirming the device extracts user intent during conversation and responds by performing the required or intended processing. (Ex. 1002 ¶142).

Indeed, Shimomura discloses analyzing voice “to execute processing required by the user” just like the ’337 patent. (*Id.* ¶143.) Shimomura describes analyzing a user’s request (e.g., for weather) and obtaining the requested information (“Today’s weather in Tokyo is sunny”). (Ex. 1004 ¶[0065].) In another example, Shimomura describes analyzing a user’s voice to execute processing (volume change). (*Id.* ¶¶[0053]-[0055].)

Thus, Shimomura discloses a process executing unit (control unit) configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user. (Ex. 1002 ¶¶141-44.) To the extent this claim element is not expressly disclosed by Shimomura, it would have been inherent. (*Id.*)

5. 1[d]: Voice-Content Generating Unit

Shimomura discloses a voice-content generating unit that generates an output sentence as text data for a voice for output to the user. (Ex. 1002 ¶¶145-47.) As discussed above, Shimomura’s control unit generates the content for output. (*Supra* §§IX.A.3-IX.A.4.) The content can be a sentence including text data like: “Today’s weather in Tokyo is sunny.” (Ex. 1004 ¶[0065]; *see also id.* ¶¶[0076], [0078]-[0079], [0092].) The control unit includes an “output management unit” and a “behavior selection control unit.” (*Id.* ¶¶[0036]-[0037], [0046]-[0047], Fig. 4; Ex. 1002 ¶146.) The output management unit “sends the character string provided by the behavior selection control unit 45 to the speech synthesis unit 48, while the speech

synthesis unit 48 generates a speech signal S4 of synthetic speech based on the supplied character string and sends it to the speaker 23” for audible presentation. (Ex. 1004 ¶[0047]; *see also id.* ¶¶ [0046], [0049]-[0050], [0063], [0083], claims 6, 14, Fig. 13.) Thus, Shimomura discloses a voice-content generating unit (control unit) configured to generate an output sentence that is text data for a voice to be output to the user. (Ex. 1002 ¶146.)

Shimomura also discloses the voice-content generating unit generates the output sentence based on content of the processing executed by the process executing unit. (*Id.* ¶147.) As discussed above, Shimomura’s control unit determines the content spoken to the user based on the recognized speech so that the device converses with the user. (*Supra* §IX.A.4.) Thus, Shimomura discloses a voice-content generating unit (control unit) configured to generate, based on content of the processing executed by the process executing unit (control unit), output sentence that is text data for a voice to be output to the user. (Ex. 1002 ¶¶145-47.)

6. 1[e]: Output Controller

Shimomura’s control unit generates the speech signal and sends it to a speaker. (*Supra* §IX.A.5.) Shimomura’s control unit also implements the speech synthesis unit to “store[] in the ROM 33 ... a volume change amount specification table 52 that specifies the amount of change in the speech volume of the robot 1 with respect to the distance to the conversation partner.” (Ex. 1004 ¶[0067]; *see also id.*

¶¶[0050], [0068]-[0073], [0075], [0083]-[0084], [0090], [0092], claims 2, 10, 18,

Figs. 9, 13.) The speech synthesis unit then

determines the reference volume read from the reference volume memory unit 49 as the current speech volume, and when the distance to the conversation partner is in the range of 50 to 80 cm, determines a volume that is a predetermined specified value (for example, ‘1’) higher than the reference volume as the current speech volume, and similarly thereafter, when the distance to the user is 80 to 150 cm, 150 to 250 cm, or 250 cm or more, determines a volume that is ‘2’, ‘3’, or ‘5’ higher than the reference volume as the current speech volume.

(*Id.* ¶[0070].) Thus, Shimomura discloses an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto.

(Ex. 1002 ¶¶148-50.)

7. 1[f]: Generate a First Output Sentence

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

1002 ¶152.) 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

(Ex. 1004, Fig. 7; Ex. 1002 ¶152.) 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.’” (Ex. 1004 ¶[0065]; Ex. 1002 ¶153.)

Thus, Shimomura discloses the voice-content generating unit is further configured to generate a first output sentence (transformed content) as the output sentence when the acquired voice has been classified as the first voice (e.g., voice farther than 350 cm). (Ex. 1002 ¶¶151-54.)

8. 1[g]: Generate a Second Output Sentence

The ’337 patent explains that omitted information may include “a polite Japanese expression such as ‘desu,’” “a Japanese postpositional particle ‘ha,’” or in

English, “an article or the like.” (Ex. 1001, 9:31-35.) This is the same type of information omitted by Shimomura’s device. (Ex. 1002 ¶156.) Shimomura discloses this element. (*Id.* ¶¶155-59.)

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]; *see also id.* ¶¶[0057]-[0058], [0065], [0085].) Thus, when the voice is classified as a second voice (e.g., closer than 350 cm), the device generates a second output sentence comprising the original content without transformation. (*Id.* ¶¶[0062], [0092]; Ex. 1002 ¶157.) 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

(Ex. 1004, Fig. 7; Ex. 1002 ¶157.) In that case, the output content lacks the added words, i.e., it omits some information provided when the voice is classified as a first voice. (Ex. 1004 ¶[0062]; Ex. 1002 ¶157.)

For example, Shimomura discloses that the output sentence when the voice is classified as the second voice (e.g., closer than 350 cm) may be: “Today’s weather in Tokyo is sunny.” (Ex. 1004 ¶[0065].) Thus, Shimomura discloses omitting some information (i.e., “ne” after phrases) when the voice is classified as the second voice. (*Id.*; Ex. 1002 ¶158.)

Thus, Shimomura discloses the voice-content generating unit is further configured to generate a second output sentence (original content) in which information is omitted (e.g., “ne”) as compared to the first output sentence (transformed content) as the output sentence when the acquired voice has been classified as the second voice (e.g., voice closer than 350 cm). (Ex. 1002 ¶¶155-59.)

9. 1[h]: Output Controller Adjusts Volume of Voice Data

Shimomura’s control unit adjusts the output volume of the content based on the user-to-device distance, as discussed for claim element 1[e]. (*Supra* §IX.A.6; Ex. 1004 ¶[0067]; Ex. 1002 ¶161.) For example, it can implement the speech synthesis unit as shown in Figure 9’s volume change table:

[FIG. 9]

Less than 50 cm	50~80cm	80~150cm	150~250cm	More than 250
0	+1	+2	+3	+5

52

FIG. 9 Volume Change Table

(*Id.*, Fig. 9, ¶[0070]; Ex. 1002 ¶161.)

As described, Shimomura discloses generating different output sentences based on distance. (*Supra* §§IX.A.7-IX.A.8; Ex. 1002 ¶162.) Although Shimomura generally discloses altering the volume based on distance, it does not specify that the output sentences have different volumes. (*Id.*) However, this modification would have been obvious to a POSITA and trivial to implement. (Ex. 1002 ¶¶162-72.)

A POSITA would have been motivated to modify Shimomura's volume change table to include a range ending at 350 cm. (*Id.* ¶163.) This would, for example, have been implemented by extending the 150-250 cm range to end at 350 cm, as shown in the highlighted column:

Less than 50 cm	50~80 cm	80~150 cm	150~350 cm	More than 350 cm
0	+1	+2	+3	+5

First Modified Volume Change Table

(*Id.*) Alternatively, this would have been implemented by adding a new range from 250 cm to 350 cm, as shown in the highlighted columns:

Less than 50 cm	50~80 cm	80~150 cm	150~250 cm	250~350 cm	More than 350 cm
0	+1	+2	+3	+5	+7

Second Modified Volume Change Table

(*Id.*) A POSITA would have been motivated to make such a modification for several reasons.

First, a modification such as the one shown in the first modified table would have been the simple substitution of a known element (a 350 cm threshold as in Shimomura's Figures 7 and 10-11) for another known element (a 250 cm threshold as in Shimomura's Figure 9) to obtain predictable results (a device that adjusts output volume based on whether user is closer or farther than 350 cm). (*Id.* ¶164); *KSR*, 550 U.S. at 417.

Second, such a modification reflects nothing more than a combination of known elements (distance thresholds disclosed in Shimomura) to achieve an expected result (adjusting output volume based on a user's distance to ensure appropriate audibility at distances around 350 cm). (Ex. 1002 ¶165); *KSR*, 550 U.S. at 417.

Third, such a modification reflects the application of a known technique (adjusting output volume with a threshold of 350 cm as in Shimomura's Figures 7 and 10-11) to achieve predictable results (adjusting output volume to ensure audibility around 350 cm). (Ex. 1002 ¶166); *KSR*, 550 U.S. at 417.

Fourth, such a modification reflects nothing more than applying a known technique (adjusting output volume with a threshold of 350 cm as in Shimomura's Figures 7 and 10-11) to a known device and method (Shimomura's control unit that varies output volume with user distance) that is ready for improvement and yields

predictable results (a device that varies output volume to ensure audibility around 350 cm). (Ex. 1002 ¶167); *KSR*, 550 U.S. at 417.

Fifth, such a modification would be a matter of routine experimentation and common sense for a POSITA. (Ex. 1002 ¶168); *KSR*, 550 U.S. at 417. Given Shimomura’s goal of enhancing user interactions by adjusting volume with distance (Ex. 1004 ¶¶[0010]-[0011], [0016]-[0017], Abstract), a POSITA would have logically modified the threshold (as in the first modified table) or added a threshold (as in the second modified table) to ensure appropriate audibility around the 350 cm threshold indicated in Shimomura’s Figures 7 and 10-11. (Ex. 1002 ¶168.) This simple modification would yield predictable improvements without requiring significant innovation. (*Id.*)

Sixth, such a modification would have been obvious to try. (*Id.* ¶169); *KSR*, 550 U.S. at 417. In environments where Shimomura’s device operates at varied distances, users may interact from beyond 250 cm. (Ex. 1002 ¶169.) Extending the “More than 250 cm” range to 350 cm, as in the first modified table, addresses this without adding discrete categories, ensuring predictable responses up to 350 cm—a logical limit aligned with Shimomura’s Figures 7 and 10–11. (*Id.*) This approach simplifies design by minimizing programming complexity. (*Id.*) Alternatively, adding a range from 250 cm to 350 cm, as in the second modified table, is an incremental adjustment continuing the trend of volume increases with distance. (*Id.*)

Shimomura's existing incremental volume changes make this a predictable solution for ensuring audibility at greater distances. (*Id.*)

A POSITA would have reasonably expected success in modifying Shimomura's volume change table in this way. (*Id.* ¶170.) This trivial modification would have involved configuring Shimomura's control unit with the revised table to provide volume levels based on whether the user is closer or farther than 350 cm. (*Id.*)

Considering the first modified volume change table and Figure 7's character string transformation determination table together, it would have been obvious to a POSITA from Shimomura that, when a voice is classified as a first voice (e.g., voice farther than 350 cm), a first output sentence containing the original content after transformation is provided at a volume level of +5 higher than the reference volume (first modified volume change table) and, when a voice is classified as a second voice (e.g., voice closer than 350 cm, such as 300 cm), a second output sentence containing original content without transformation may be provided at a volume level of +3 higher than the reference volume (first modified volume change table). (*Id.* ¶171.)

			Second Voice	First Voice
Less than 50 cm	50~80 cm	80~150 cm	150~350 cm	More than 350 cm
0	+1	+2	+3	+5

First Modified Volume Change Table

(Id.)

[FIG. 7]

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

FIG. 7 Character string transformation determination table

(Ex. 1004, Fig. 7; Ex. 1002 ¶171.) Similarly, considering the second modified volume change table and Figure 7's character string transformation determination table together, it would have been obvious to a POSITA from Shimomura that, when a voice is classified as a first voice (e.g., voice farther than 350 cm), a first output sentence containing the original content after transformation is provided at a volume level of +7 higher than the reference volume (second modified volume change table) and, when a voice is classified as a second voice (e.g., voice closer than 350 cm,

such as 300 cm), a second output sentence containing original content without transformation may be provided at a volume level of +5 higher than the reference volume (second modified volume change table). (Ex. 1002 ¶171.)

				Second Voice	First Voice
Less than 50 cm	50~80 cm	80~150 cm	150~250 cm	250~350 cm	More than 350 cm
0	+1	+2	+3	+5	+7

Second Modified Volume Change Table

(*Id.*)

[FIG. 7]

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

FIG. 7 Character string transformation determination table

(Ex. 1004, Fig. 7; Ex. 1002 ¶171.) Shimomura expressly discloses that both the content and volume may vary. (Ex. 1004 ¶¶[0085] (“[T]he robot 1 speaks louder ... the farther away the user/conversation partner is, while depending on the distance to the user, it ... inserts the word ‘ne’ after each phrase.”), [0083]; *see also id.* ¶¶[0084], [0092], Fig. 4, claims 6, 14, 21; Ex. 1002 ¶171.)

Thus, it would have been obvious to a POSITA to modify Shimomura’s output controller (control unit) to be configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence (transformed content) thereinto differs from the sound volume of the voice data obtained by converting the second output sentence (original content) thereinto. (Ex. 1002 ¶¶160-72.)

Accordingly, Shimomura renders obvious claim 1. (*Id.* ¶¶130-73.)

B. Claim 2

Shimomura discloses the additional limitations of claim 2. (Ex. 1002 ¶¶174-85.)

1. 2[a]: Process Executing Unit

Shimomura’s control unit determines the output content by extracting and responding to a user’s intention, as discussed above. (*Supra* §IX.A.4.) Shimomura discloses that the device controls its speech “during a dialogue between the device and a user,” enabling the device to “converse with the user smoothly.” (Ex. 1004 ¶¶[0010], [0086].) Shimomura’s control unit acquires content such as “Today’s weather in Tokyo is sunny” for presenting to the user. (*Id.* ¶¶[0044], [0065]; Ex. 1002 ¶176.)

Even if Shimomura did not expressly disclose this element, Shimomura’s control unit inherently includes an “intention analyzing unit” and an “acquisition content

information acquiring unit,” as recited in claim element 2[a], because these were necessary for Shimomura’s device to function as described. (Ex. 1002 ¶177.) These functionalities are essential for interpreting user input and generating appropriate responses. (*Id.*) 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 “acquisition content information acquiring unit.” (*Id.*) These units are required for Shimomura’s control unit to perform as disclosed. (*Id.*)

Thus, Shimomura expressly or inherently discloses the process executing unit (control unit) comprises: an intention analyzing unit configured to extract intention information indicating an intention of the user based on the voice acquired by the voice acquiring unit. (*Id.* ¶¶175-78.) Shimomura also expressly or inherently discloses that the process executing unit (control unit) comprises an acquisition content information acquiring unit configured to acquire acquisition content information which is notified to the user based on the extracted intention information. (*Id.*)

Even if Shimomura had not disclosed this claim element (expressly or inherently), the claim element would have been obvious to a POSITA from Shimomura. (*Id.* ¶179.) A POSITA would have been motivated to use an “intention analyzing

unit” to determine user intent and an “acquisition content information acquiring unit” to retrieve content based on that intent for several reasons. (*Id.*)

Shimomura describes voice processing capabilities that allow the device to interpret user intent and generate appropriate responses. (*Supra* §IX.A.4.) A POSITA would have been motivated to structure Shimomura’s control unit with an intention analyzing unit and a content acquisition unit to perform these functions more efficiently. (Ex. 1002 ¶180.) 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 voice analysis systems worked at the time of the invention. (*Id.*; *see infra* §XI.)

This approach would have been straightforward for a POSITA, as Shimomura’s device already integrates voice analysis and content generation. (Ex. 1002 ¶181.) Implementing these functionalities as units within Shimomura’s control unit would have been an obvious design choice, enhancing modularity, maintainability, and troubleshooting. (*Id.*) This design would predictably improve response accuracy, while offering greater flexibility to handle various user requests. (*Id.*)

Thus, this claim element would have been obvious over Shimomura. (*Id.* ¶¶175-82.)

Moreover, as discussed in Ground 1, even if Shimomura had not disclosed or suggested this claim element, it would not make the claim patentable because the '337 patent admits this claim element was known in the art. (*Supra* §VII.B.1.)

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

Shimomura's control unit acquires information, such as weather information, to determine the content spoken to the user. (*Supra* §IX.B.1; Ex. 1004 ¶¶[0010], [0065], [0086].) This information is incorporated into the voice content provided to the user, as demonstrated by the example discussed for claim element 2[a]. (*Id.*; Ex. 1002 ¶184.) The control unit's output management unit "sends the character string provided by the behavior selection control unit 45 to the speech synthesis unit 48, while the speech synthesis unit 48 generates a speech signal S4 of synthetic speech based on the supplied character string and sends it to the speaker 23." (Ex. 1004 ¶[0047].) Thus, Shimomura discloses that the voice-content generating unit (control unit) is configured to generate the text data including the acquisition content information as the output sentence. (Ex. 1002 ¶184.)

Accordingly, Shimomura renders obvious claim 2. (*Id.* ¶¶174-85.)

C. Claim 3

Shimomura's control unit generates content, such as "Today's weather in Tokyo is sunny," while conversing with the user. (*Supra* §§IX.A.4, IX.A.8.) A POSITA would have understood that this content could be prompted by specific user

speech, such as by asking, “what is the weather today?” (Ex. 1002 ¶187.) The interrogative word prompts processing, and the device omits it in the response to provide a direct answer rather than echoing the question. (*Id.*)

A POSITA also would have been motivated for several reasons to configure Shimomura’s control unit to omit some user-spoken information when responding. (*Id.* ¶¶188-92.)

First, this would have been a predictable and desirable implementation of Shimomura’s control unit, which processes and responds to user inputs. (*Id.* ¶189); *KSR*, 550 U.S. at 417. A POSITA would have understood that repeating all user-spoken information is unnecessary, and excessive repetition would be inefficient and delay responses. (Ex. 1002 ¶189.)

Second, market demands and user expectations for efficient, streamlined interactions with automated systems would have motivated this implementation. (*Id.* ¶190); *KSR*, 550 U.S. at 417. Users expect minimal delays and concise responses. (Ex. 1002 ¶190.) Omitting user-spoken information would align Shimomura’s control unit with these expectations, improving user satisfaction. (*Id.*) A POSITA would have understood this as an obvious solution. (*Id.*)

Third, such an implementation would have been a matter of routine experimentation and common sense for a POSITA. (*Id.* ¶191); *KSR*, 550 U.S. at 417. Given Shimomura’s aims to enhance user interactions (Ex. 1004 ¶¶[0010]-[0011],

[0016]-[0017], Abstract), reducing unnecessary repetition would have been a logical step. (Ex. 1002 ¶191.) A POSITA would have recognized this simple, routine modification (if even a modification) would yield predictable improvements without requiring significant innovation. (*Id.*)

Fourth, such an implementation would have been obvious to try. (*Id.* ¶192); *KSR*, 550 U.S. at 417. Shimomura’s device presents only two design options for responding to a user request: include all information from a user’s request or less than all. (Ex. 1002 ¶192.) Omitting some information to enhance efficiency and conciseness would have been an obvious approach. (*Id.*)

A POSITA would have reasonably expected success in implementing Shimomura’s control unit to omit some user-spoken information when responding. (*Id.* ¶193.) This implementation would have been trivial, involving straightforward changes to configure the control unit for more efficient and streamlined interactions. (*Id.*)

Thus, Shimomura renders obvious that, on generating the second output sentence, the voice-content generating unit (control unit) is further configured to omit a part of information included in the voice spoken by the user and thus renders obvious claim 3. (*Id.* ¶¶186-95.)

D. Claims 4 and 5

Shimomura discloses both a speech control device and method. (Ex. 1004 ¶¶[0012]-[0013]; Ex. 1002 ¶196; *see also* Ex. 1004 ¶¶[0001], [0009], [0016], [0057], Title, Abstract, Figs. 6, 8, 13, claims 9-16.) Shimomura also discloses electronic devices implemented by instructions stored in a computer-readable storage media. (Ex. 1004 ¶¶ [0027], [0030], [0036]; Ex. 1002 ¶196; *see also* Ex. 1004 ¶¶[0028]-[0029], [0037]-[0047], [0051]-[0083], Figs. 3-4.)

As discussed above, the remainder of claims 4 and 5 are identical and merely recite the functions performed by the various units recited in claim 1. (*Supra* §VII.D; Ex. 1002 ¶197.) Thus, for the reasons provided for claim 1, Shimomura renders obvious claims 4 and 5. (*Id.*)

X. GROUND 4: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIMOMURA AND SHIN.

A. Claim 1

Shimomura and Shin individually disclose and render obvious all limitations of claim 1 as discussed for Grounds 1-3. (*Supra* §§VII.A, VIII.A, IX.A.) Even if Shimomura alone did not teach or motivate a POSITA to alter the output's content and volume at the same threshold distance (e.g., 350 cm), doing so would have been obvious in view of Shin. (Ex. 1002 ¶¶199-204.)

Both Shimomura and Shin teach proximity-based voice output adjustments for improving user interaction. (*Id.* ¶200.) A POSITA would have naturally combined these teachings to enhance adaptability and functionality in voice-controlled devices with predictable results. (*Id.*)

Shin's volume change table (Table 1) and amount of information provided table (Table 3) include matching distance ranges, correlating specific distances with volume levels and information amounts. (*Id.* ¶201.) For example, Shin uses Tables 1 and 3 to classify the user's voice into a first category (e.g., first voice) based on a user distance of less than one meter, and a second category (e.g., second voice) based on a user distance between one to two meters:

TABLE 1

	Distance (D) between user and electronic device	Output volume level of content
First Voice	$D < 1 \text{ m}$	40 dB
Second Voice	$1 \text{ m} \leq D < 2 \text{ m}$	45 dB
	...	...
	$D \geq 5 \text{ m}$	65 dB

TABLE 3

	Distance (D) between user and electronic device	Output amount of information of content
First Voice	$D < 1 \text{ m}$	Amount of information of original content
Second 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

(Ex. 1003, Tables 1, 3; Ex. 1002 ¶201.) This would have motivated a POSITA and provided a reasonable expectation of success in modifying Shimomura’s volume change table—including in the changes discussed in Ground 3—to implement common distance thresholds for adjusting volume level and content. (Ex. 1002 ¶201.) This would enable Shimomura’s device to classify a voice into one of two distance-based categories and adjust output volume level and content based on that category, rather than referencing tables with distinct ranges. (*Id.*)

Shin discloses that the distances and volume levels in its volume change table (Table 1), similar to Shimomura’s, are “merely provided as an example” and that embodiments “are not limited to the example of Table 1.” (Ex. 1003 ¶[0081]; *see also id.* ¶[0090].) This would have motivated a POSITA and provided a reasonable expectation of success in implementing alternative distances or volume levels in Shimomura’s volume change table, such as those discussed in Ground 3, to achieve desired volume levels for specific distances. (Ex. 1002 ¶202.)

Moreover, Shin discloses that information may be classified “dichotomously” (i.e., two categories) or divided “into several levels.” (Ex. 1003 ¶[0051].) A POSITA would have been motivated to use two different volume levels for the dichotomous embodiment, with one corresponding to each category of content, aligning with Shin’s disclosure, simplifying design, and being trivial to implement. (Ex. 1002 ¶203.)

Accordingly, Shimomura and Shin render obvious claim 1. (*Id.* ¶¶199-204.)

B. Claim 2

Shimomura and Shin individually disclose and render obvious all limitations of claim 2 as discussed for Grounds 1-3. (*Supra* §§VII.B, VIII.B, IX.B.) Claim 2 also would have been obvious over Shimomura in view of Shin. (Ex. 1002 ¶205.) Specifically, Shin discloses that its processor analyzes the requests from the voice input to identify the content the user is interested in, and then acquires and provides that content. (*Supra* §§VII.A.4, VII.B.1.) Shimomura similarly describes this functionality. (*Supra* §§IX.A.4, IX.B.1.) A POSITA would have been motivated by Shin’s disclosure to implement an “intention analyzing unit” to extract intention information from user voice data and an “acquisition content information acquiring unit” to retrieve content based on the extracted intention, in Shimomura’s device. (Ex. 1002 ¶205); *KSR*, 550 U.S. at 417. Shin’s voice-processing capabilities, such as interpreting user intent for requests (e.g., time or weather), would have motivated structuring Shimomura’s processor with these units for efficient intent extraction and content retrieval. (*Supra* §§VIII.B, IX.B.) This modular design would enhance clarity, reliability, and efficiency. (Ex. 1002 ¶205.)

C. Claim 3

Shimomura and Shin individually disclose and render obvious all limitations of claim 3 as discussed for Grounds 1-3. (*Supra* §§VII.C, VIII.C, IX.C.) Claim 3

would also have been obvious over Shimomura in view of Shin. (*Supra* §VII.C; Ex. 1002 ¶¶206-14.) A POSITA would have been motivated, for several reasons, to implement Shin’s feature of omitting some user-spoken information into Shimomura’s control unit. (Ex. 1002 ¶¶206-12.)

First, such an implementation would represent merely the simple substitution of a known element (a response that omits user-spoken information, as in Shin) for another known element (Shimomura’s response) to obtain predictable results (more efficient and concise user interaction). (*Id.* ¶¶207-12); *KSR*, 550 U.S. at 417. Second, such an implementation reflects merely a combination of known elements (Shimomura’s control unit and Shin’s output that omits some user-spoken information) to achieve an expected result (efficient and streamlined interactions). (*Id.*) Third, such an implementation reflects the application of a known technique (Shin’s omission of some user-spoken information) to achieve predictable results (efficient communication with Shimomura’s device). (*Id.*) Fourth, such an implementation reflects applying a known technique (omitting some user-spoken information) to a known device and method (Shimomura’s control unit) that is ready for improvement and yields predictable results (more efficient and streamlined interaction). (*Id.*)

Fifth, such an implementation would be a straightforward optimization of Shimomura’s control unit. (*Id.*) A POSITA would have recognized Shin’s goal of enhancing user interaction efficiency by eliminating redundant information. (Ex.

1002 ¶211.) Applying this to Shimomura’s control unit would involve minimal modification (if any), yielding predictable benefits like faster communication and more concise responses. (*Id.*)

Sixth, such an implementation would have been obvious to try. (Ex. 1002 ¶212); *KSR*, 550 U.S. at 417. Shimomura’s device presents only two design options for responding to requests: include all information from a user’s request or less than all. (*Supra* §IX.C.) Given Shin’s disclosure of omitting information from user requests to enhance efficiency (*supra* §VIII.C), it would have been obvious to apply a similar approach with Shimomura’s device. (Ex. 1002 ¶212.)

A POSITA would have reasonably expected success in implementing Shin’s feature. (*Id.* ¶213.) It would have been trivial, involving straightforward changes to enhance the control unit’s efficiency and streamline user interactions. (*Id.*)

Accordingly, Shimomura and Shin render obvious claim 3. (*Id.* ¶¶206-14.)

D. Claims 4 and 5

Shimomura and Shin individually disclose and render obvious all limitations of claims 4 and 5 for the reasons discussed for Grounds 1-3. (*Supra* §§VII.D, VIII.D, IX.D.) Claims 4-5 are also obvious over Shimomura in view of Shin for the reasons explained regarding claims 4-5 in Ground 3 and claim 1 in this Ground. (Ex. 1002 ¶215; *supra* §§IX.D, X.A.)

XI. GROUND 5: CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS OVER SHIN AND/OR SHIMOMURA IN VIEW OF KRISTJANSSON.

As discussed above, Shin and/or Shimomura disclose or render obvious each limitation of claims 1-5. (*Supra* §§VII-X.) The claims also would have been obvious in view of those references and in further view of Kristjansson. Specifically, to the extent PO argues that Shin and/or Shimomura do not disclose the “process executing unit” limitations (e.g., claim limitations 1[c], 2[a], 4[d], and 5[d]), Kristjansson discloses those limitations and renders the claims obvious. (Ex. 1002 ¶¶216-23.)

For example, claim element 2[a] recites that the “process executing unit” comprises an “intention analyzing unit” and an “acquisition content information acquiring unit.” These limitations refer to determining the user’s intent using natural language understanding (“NLU”) processing and acquiring the desired information. (Ex. 1001, 4:31-33; Ex. 1002 ¶217.)

Kristjansson discloses a “speech processing system” that, similar to those disclosed in Shin and Shimomura, comprises “one or more speech-controlled devices 110 local to user 5.” (Ex. 1013, 2:45-67.) Kristjansson’s system is “capable of performing traditional speech processing” and “may execute certain commands, such as answering spoken utterances of the users[.]” (*Id.*) Kristjansson discloses that to do so, the system includes “an NLU component” that “takes textual input (such as

processed from ASR [automatic speech recognition] 450 based on the utterance input audio 11 [from the user]) and ... derive[s] an intent or a desired action from the user[.]” (*Id.*, 7:26-67.) Upon determining the user’s intent, the system sends that information to a “command processor 490,” which retrieves content based on that intention. (*Id.*, 10:38-11:12.) For example, the command processor may retrieve a song requested by the user (*id.*, 10:44-48), results of a user’s search (*id.*, 10:49-56), or content from another source requested by the user (*id.*, 10:56-11:12). In each case, the retrieved content is then delivered to the user. (*Id.*, 10:38-11:12.) Thus, Kristjansson discloses limitation 2[a]. (Ex. 1002 ¶218.) For the same reasons, Kristjansson discloses the broader limitations in the other claims, i.e., claim elements 1[c], 4[d], and 5[d]. (*Id.*)

As discussed above, Shin and Shimomura, like Kristjansson, disclose devices that receive voice input from a user and return voice output. (*Supra* §§VII-X.) A POSITA would have been motivated to implement Kristjansson’s NLU teachings in Shin and/or Shimomura’s devices to process the user’s input and determine and retrieve the proper output for several reasons. (Ex. 1002 ¶¶219-22.)

First, Shin and Shimomura suggest doing so. (*Id.* ¶220.) Shin and Shimomura both disclose systems that analyze voice input to identify the content the user is interested in, and then generate the requested content. (*Supra* §§VII-X.) Kristjans-

son's disclosure merely provides additional detail about how a POSITA would structure such a system, with NLU processing and content retrieval. (Ex. 1002 ¶220.) Thus, a POSITA would have been motivated to look to Kristjansson when implementing Shin and/or Shimomura's systems. (*Id.*)

Second, Kristjansson's disclosure represents a routine and well-established approach to providing a user with requested content. (*Id.* ¶221; *see* Ex. 1015, 431-33); *KSR*, 550 U.S. at 417. Third, such a modification reflects merely applying a known technique (Kristjansson's NLU processing) to achieve predictable results (providing a user with requested content). (*Id.*) Fourth, such a modification reflects merely applying a known technique (Kristjansson's NLU processing) to a known device (Shin's and/or Shimomura's) that is ready for improvement and yields predictable results (providing a user with requested content). (*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 ¶222) and because Shin and Shimomura contemplate modification or addition to their systems by POSITAs (*e.g.*, Ex. 1003 ¶[0184]; Ex. 1004 ¶[0089]; Ex. 1002 ¶222.) Moreover, Kristjansson discloses a system, like Shin and Shimomura, that uses a proximity sensor to determine the user-to-device distance and adjusts the volume of the output based on that distance. (Ex. 1013, claim 1, 3:1-27; Ex. 1002 ¶222.)

Accordingly, Shin and/or Shimomura, in further view of Kristjansson, render obvious claims 1-5. (Ex. 1002 ¶¶216-23.)

XII. 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.

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

The Board should not exercise its discretion under §314(a). PO asserted the '337 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. The Petition Presents Compelling Evidence of Unpatentability.

The Board “will not rely on the *Fintiv* factors to 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. Shin, Shimomura, and Kristjansson plainly anticipate and render obvious the challenged claims.

Fairness also warrants institution here, 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 interest also supports institution. Thus, in view of the compelling evidence of unpatentability, Petitioner’s stipulation, fairness, and for the public’s benefit, the Board should not discretionarily deny this petition.

XIV. 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).

Advanced Bionics Prong 1. Shimomura, which is relied on in Grounds 3-5 to challenge all claims, was never before the Examiner. 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). Kristjansson was likewise not before the Examiner.

Furthermore, although the Examiner identified Shin, the Examiner did not apply it in any rejection. Instead, the Examiner cited Shin as “pertinent” for its

disclosure in Paragraph [0037] that a voice output response could be “adjusted or determined *according to the volume of the user input[.]*” (Ex. 1007, 62-63 (emphasis added).) The Examiner clearly overlooked the most relevant disclosures in Shin, including that the device’s voice response outputs *different content* (*supra* §§VII.A.7-VII.A.8) at a *different volume* (*supra* §VII.A.9) based on *the distance between the user and the device* (*supra* §§VII.A.7-VII.A.9). These are the very elements that the Examiner believed were *not* disclosed and which the Examiner relied on in allowing the claims. The Examiner’s two-line citation of Shin, which does not discuss Shin’s most pertinent disclosures, shows the Examiner failed to meaningfully address Shin. Under such circumstances, the Board has declined to exercise discretion under the first prong. *JUUL Labs, Inc. v. NJOY, LLC*, IPR2024-00160, Paper 10 at 11-13 (P.T.A.B. May 24, 2024); *Samsung Elecs. Co. v. Maxell, Ltd.*, IPR2024-00867, Paper 9 at 17, 20 (P.T.A.B. Nov. 7, 2024).

Because the first prong is unmet, the Board “need not address the second” prong. *JUUL*, Paper 10 at 13; *Quasar Sci. LLC v. Colt Int’l Clothing, Inc.*, IPR2023-00611, Paper 10 at 12 (P.T.A.B. Oct. 10, 2023).

Advanced Bionics Prong 2. Should the Board reach the second prong, it too favors institution. The Examiner did not consider Shimomura or Kristjansson, nor apply Shin. 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 other than an Examiner ... listing a reference on a Notice of References Cited, 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 (“[I]f the record of the Office’s previous consideration of the art is not well developed ... then 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). As discussed, the Examiner’s citation to Shin failed to discuss highly relevant disclosure presented in this Petition and failed to reject the claims over that disclosure, representing material error. Thus, the Board should not deny institution under §325(d).

XV. 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 ’337 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 BoxSEAZN2L2104LP2@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 BoxSEAZN2L2104LP2@knobbe.com.

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

Petitioner certifies that the '337 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 five claims is requested. Payment for any additional fees due may be charged to the above-referenced Deposit Account.

Dated: January 31, 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. 11,069,337	
Claim 1	
1[pre]	A voice-content control device, comprising:
1[a]	a proximity sensor configured to calculate a distance between a user and the voice-content control device;
1[b]	a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device;
1[c]	a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user;
1[d]	a voice-content generating unit configured to generate, based on content of the processing executed by the process executing unit, output sentence that is text data for a voice to be output to the user; and
1[e]	an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto, wherein
1[f]	the voice-content generating unit is further configured to generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice, and generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice, wherein
1[g]	
1[h]	the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto.

Listing of Claims from U.S. 11,069,337	
Claim 2	
2[pre]	The voice-content control device according to claim 1, wherein
2[a]	the process executing unit comprises: an intention analyzing unit configured to extract intention information indicating an intention of the user based on the voice acquired by the voice acquiring unit; and an acquisition content information acquiring unit configured to acquire acquisition content information which is notified to the user based on the extracted intention information, and
2[b]	the voice-content generating unit is further configured to generate the text data including the acquisition content information as the output sentence.
Claim 3	
--	The voice-content control device according to claim 1, wherein, on generating the second sentence, the voice-content generating unit is further configured to omit a part of information included in the voice spoken by the user.

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

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

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. 11,069,337** contains 13,945 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: January 31, 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. 11,069,337**
and **EXHIBITS 1001-1015**, are being served on January 31, 2025 via Federal Ex-
press overnight mail on counsel of record for U.S. Patent No. 11,069,337, 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: January 31, 2025

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