

Petitioners' Demonstratives

May 20, 2026

Amazon.com, Inc. v. SoundClear Technologies, LLC

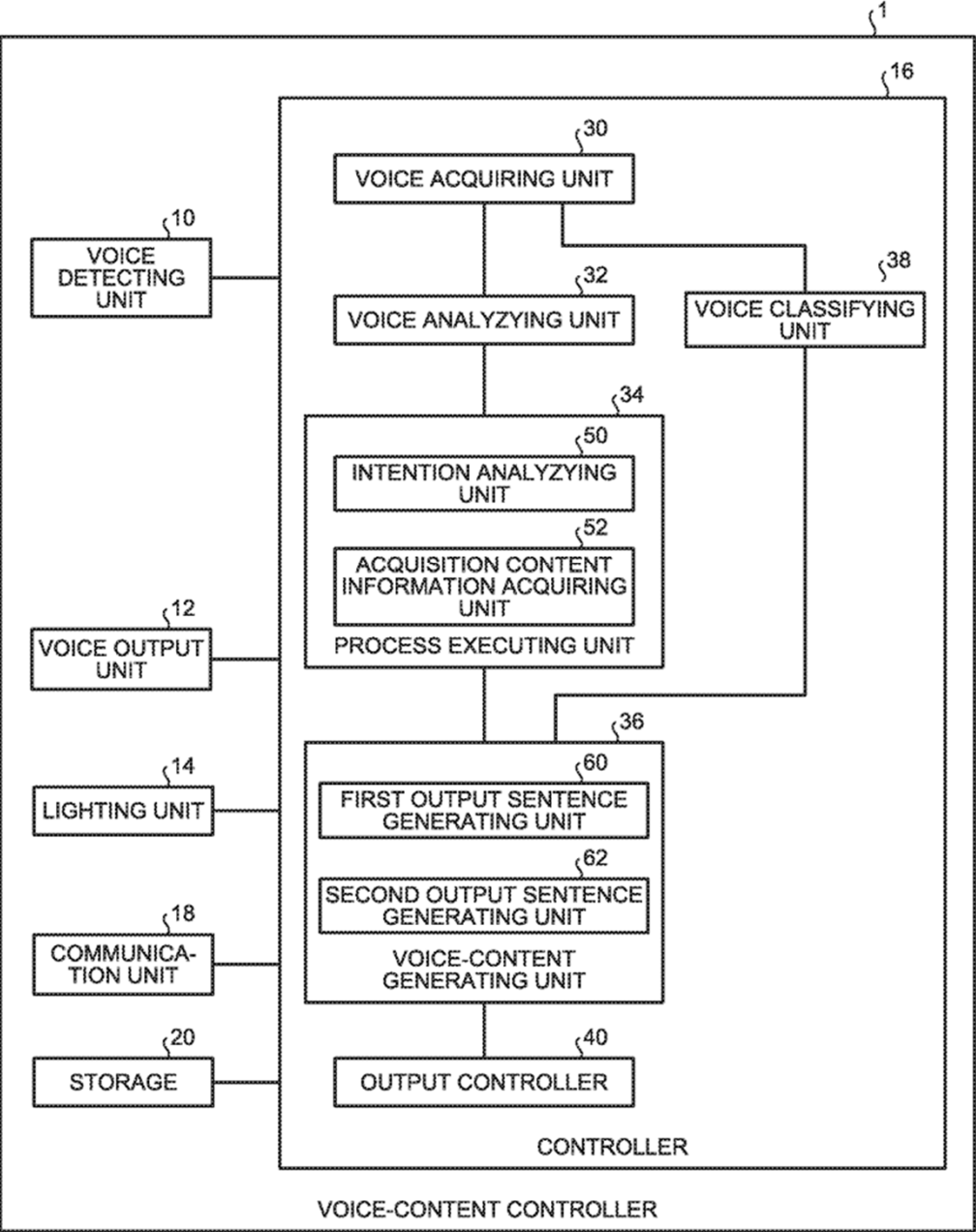
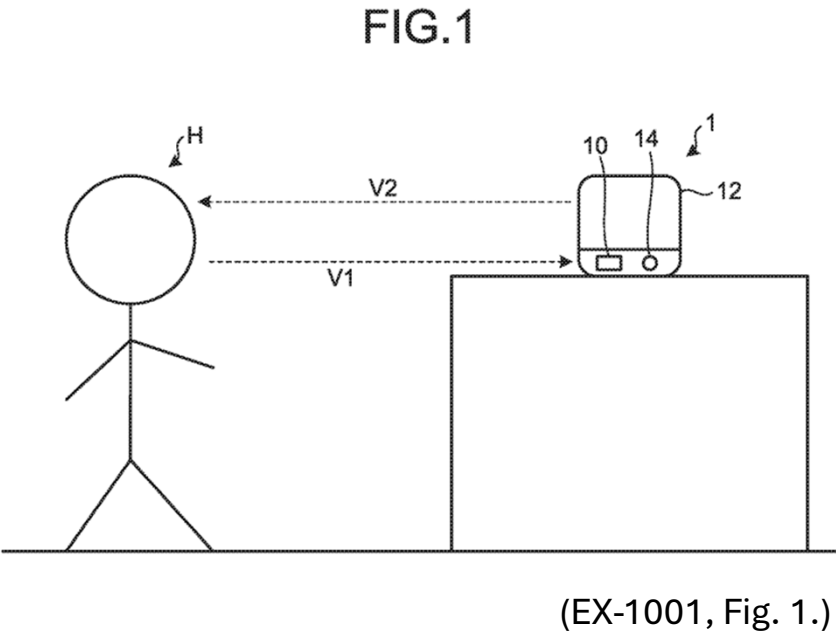
IPR2025-00565
U.S. Patent No. 11,069,337



Amazon v. SoundClear
US Patent 11,069,337
Amazon EX-1301

'337 Patent Figures

FIG.2



(EX-1001, Fig. 2.)

Claim 1

1. A voice-content control device, comprising:

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

[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;

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

[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

[e] **an output controller** configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto, wherein the voice-content generating unit is further configured to

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

[g] *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

[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.

Claim 1

1. A voice-content control device, comprising:

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

[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;

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

[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

[e] **an output controller** configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto, wherein the voice-content generating unit is further configured to

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

[g] *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

[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.

Claim 1

1. A voice-content control device, comprising:

[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;**

...

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

[g] generate a **second output sentence in which information is omitted** . . . when the acquired voice has been classified as the **second voice**, wherein

[h] . . . adjust the sound volume of voice data such that the sound volume of . . . the first output sentence . . . differs from the sound volume of the . . . second output sentence . . .

Claim 1: Key Limitations

1. Classify voice based on distance
2. Output:
 - First Voice → full output
 - Second Voice → omit some information
3. Different volume for each output

Relevant Disclosure in '337 Patent

10 threshold. Note that the voice classifying unit 38 can perform the classification to the first voice V1A and the second voice V1B by using any method. For example, the voice classifying unit 38 can perform the classification to the first voice V1A and the second voice V1B by using a slope of the
15 peak in the spectrum as a feature value. Moreover, the voice classifying unit 38 can perform the classification to the first voice V1A and the second voice V1B by using either one of a volume of the voice V1, a speaking speed of the user in the voice V1, and a volume ratio between a speech of the user
20 and a wind noise as a feature value. Furthermore, a proximity sensor can be provided in the voice-content control device 1, a distance between the user H and the voice-content control device 1 is calculated from a detection result of the proximity sensor, and the distance can be used as a
25 feature value to perform the classification to the first voice V1A and the second voice V1B. Moreover, the voice classifying unit 38 can derive a Mel-frequency cepstrum coefficient as a feature value, and can perform the classification to the first voice V1A and the second voice V1B based on
30 the mel-frequency cepstrum coefficient. In these cases, the voice classifying unit 38 sets a threshold of the feature value, and classifies the voice V1 as either one of the first voice V1A or the second voice V1B based on whether the feature value exceeds the threshold.

(EX-1001, 8:10-34.)

First Voice: Feature Value (D) ≤ threshold (x)
Second Voice: Feature Value (D) > threshold (x)

'337 Patent Example

User: “What’s the weather today?”

First output sentence (first voice):

“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[.]”

Second output sentence (second voice):

~~“On the day Z of the month Y of the year X, in Tokyo,]~~ [T]he 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.”

Volume:

“[T]he output controller 40 may use different sound volumes ... between the first output sentence and the second output sentence.”

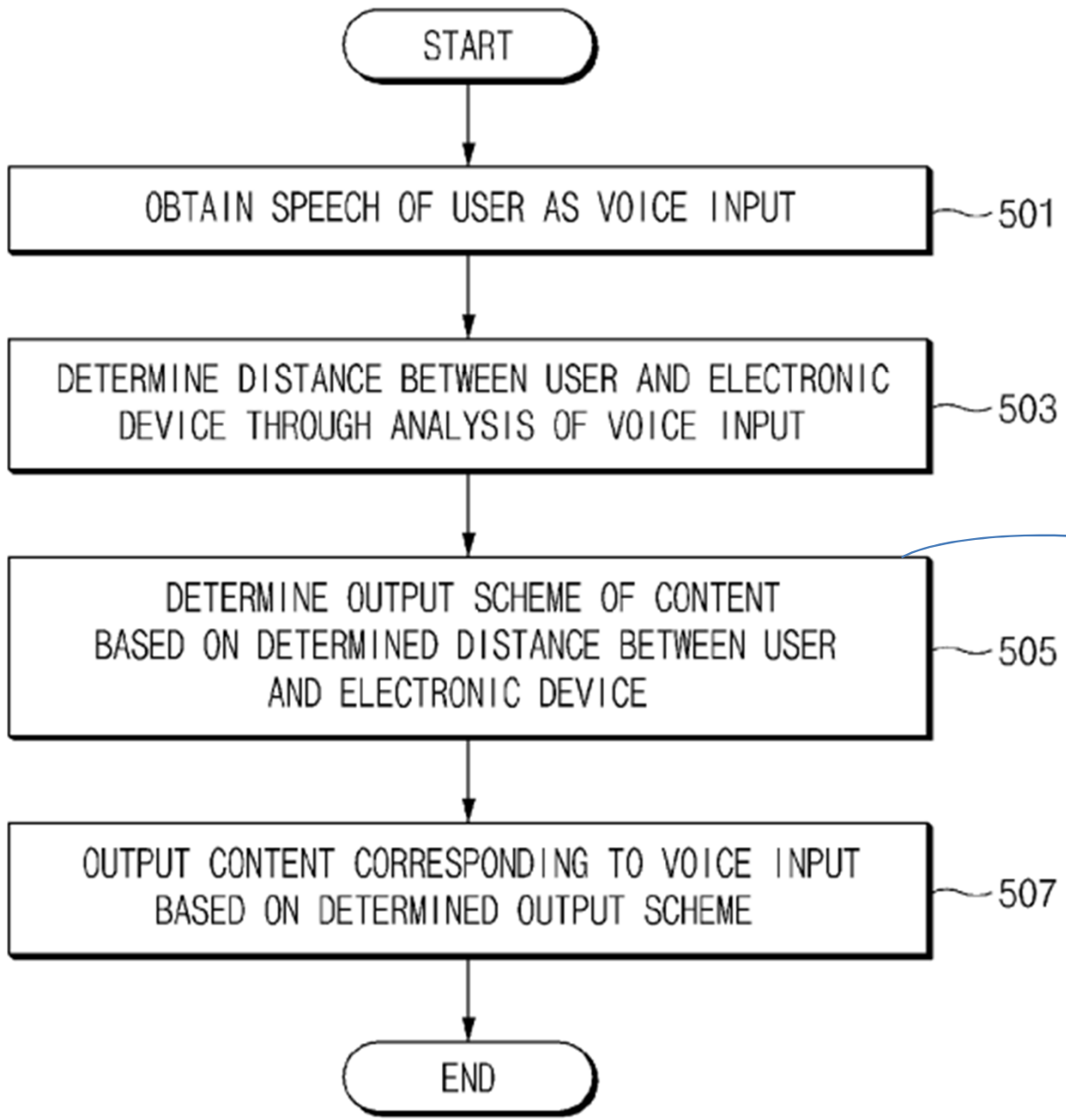
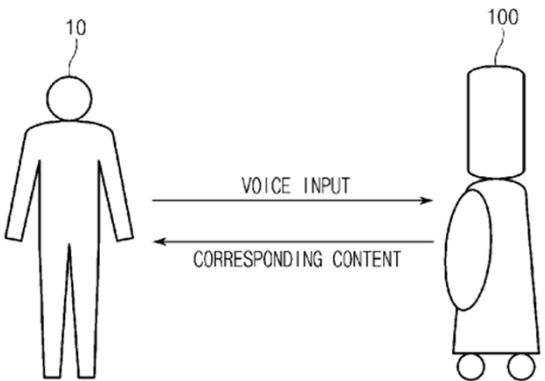
(EX-1001, 4:56-61, 9:16-21 (first sentence), 10:18-25 (second sentence), 13:62-65 (volume).)

Grounds

Ground	Claims	Basis	References
1	1-5	§ 102	Shin
2	1-5	§ 103	Shin
3	1-5	§ 103	Shimomura
4	1-5	§ 103	Shimomura + Shin
5	1-5	§ 103	Shin and/or Shimomura + Kristjansson

Shin

Shin Overview



“output scheme”
=
“output volume level ... and an
output amount of information”
(EX-1003 ¶[0049].)

(EX-1003, Fig. 5A.)

Shin's Disclosure

User: “Let me know today’s weather.”

First output sentence (first voice):

“The weather in Jul. 1 2015 is (①) rainy after cloudy (②). The highest temperature 28°C., and the lowest temperature is 18 °C (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).”

Second output sentence (second voice):

~~“The weather in Jul. 1 2015 is (①) rainy after cloudy (②). The highest temperature 28°C., and the lowest temperature is 18 °C (③), the rainfall is 10 mm (④). Prepare your umbrella when you go out. (⑤).”~~

Volume:

The “output volume level of the corresponding content increases as a distance between the user and the electronic device 101 increases[.]”

(EX-1003 ¶¶[0088] (output sentences), [0079] (volume).)

Shin Discloses Classifying the Voice Based on Distance

TABLE 3

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

(EX-1003, Table 3 (consolidated).)

Shin Discloses Classifying the Voice Based on Distance

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
	$1\text{ m} \leq D < 2\text{ m}$	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
Second Voice	$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

(EX-1003, Table 3 (annotated); EX-1002 ¶64.)

“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[.]”

(EX-1002 ¶64.)

Shin Discloses Generating the Claimed Output Sentences

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)

(EX-1003, Table 3 (annotated); EX-1002 ¶81.)

“[W]hen the voice is classified as a second voice (e.g., greater than 4 meters), the output content is abbreviated, i.e., it omits some information provided when the voice is classified as a first voice.”

(EX-1002 ¶81.)

Shin Discloses Outputting the Sentences at Different Volumes

The “output volume level of the corresponding content increases as a distance between the user and the electronic device 101 increases[.]”

(EX-1003 ¶[0079]; *id.* ¶[0080].)

TABLE 1

Distance (D) between user and electronic device	Output volume level of content
D < 1 m	40 dB
1 m ≤ D < 2 m	45 dB
...	...
D ≥ 5 m	65 dB

(EX-1003, Table 1.)

Shin Discloses the Disputed Limitations

TABLE 3		TABLE 1	
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)	

Distance (D) between user and electronic device		Output volume level of content	
D < 1 m		40 dB	
1 m ≤ D < 2 m		45 dB	
...		...	
D ≥ 5 m		65 dB	

(EX-1003, Table 3 (annotated); EX-1002 ¶81.)

(EX-1003, Table 1.)

“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). When a voice is classified as a second voice (e.g., greater than 4 meters), a second output sentence containing ‘[c]ontent of fourth abbreviation level’ is provided (Table 3) at a volume level of 60 or 65 dB (Table 1 and ¶[0080].)”

(EX-1002 ¶88.)

The Board's Preliminary Claim Construction Was Correct

Construction: “user” means “one or more users”

(DI, 15.)

PO's expert agrees:

7 Q. You agree that this claim only requires one
8 user speaking to the device, correct?
9 MR. SPRENGER: Objection; vague, and
10 to the extent it calls for a legal conclusion.
11 A. The claim requires -- you're referring to
12 the phrase "a user"? Is that what you --
13 BY MR. HEIDEMAN:
14 Q. Let's start with that.
15 A. All right. "A user" means one or more. So
16 it's not limited to one user. It could be more than
17 one.
18 Q. Okay. It could be just one person speaking
19 to the device, right?
20 A. It could be, sure.
21 Q. And that person's voice is classified into
22 one of two categories, right?
23 A. Correct.

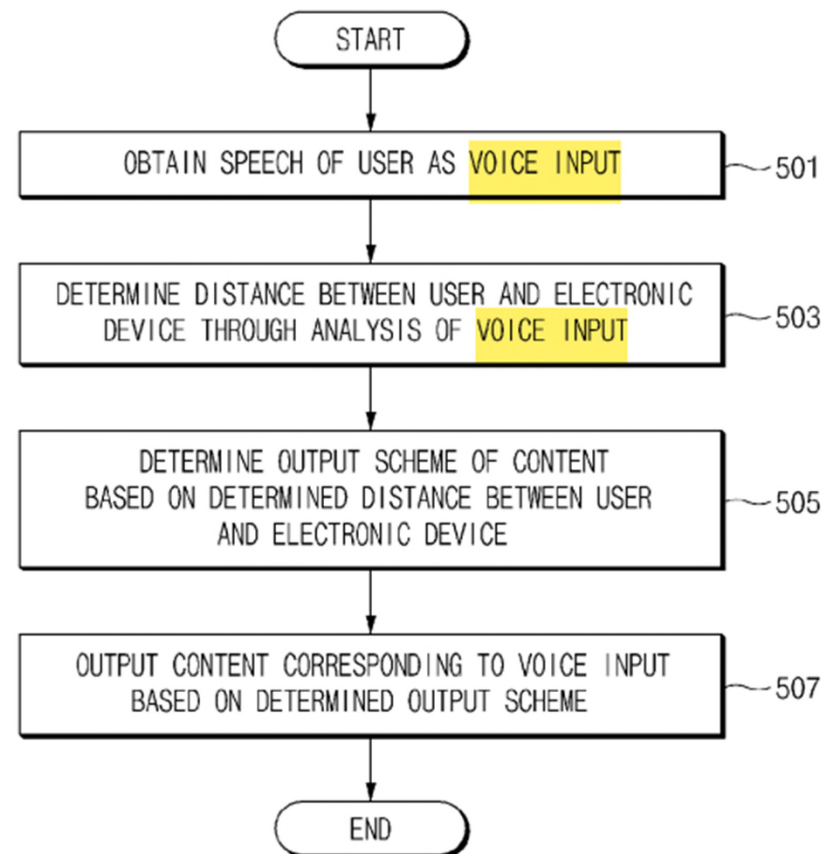
(EX-1201, 30:7-23.)

PO Argument 1 Fails

PO's Argument 1: Shin only mentions “voice” in connection with the output. (POR, 24.)

Shin's Disclosure: the device “receive[s] a speech of a user as a **voice input**”

(EX-1003, Abstract.)



(EX-1003, Fig. 5A; EX-1202 ¶17.)

PO Argument 2 Fails

PO’s Argument 2: “Table 3 discloses only a calculation of distance and nothing else”
(POR, 25.)

Shin’s Disclosure:

TABLE 3	
Distance (D) between user and electronic device	Output amount of information of content
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)

(EX-1003, Table 3.)

PO Argument 3 Fails

PO's Argument 3: Shin's device would not distinguish between the voices of two different users standing the same distance from the device. (POR, 25.)

Inconsistent with Claim Language: The claim does not require different users' voices to be classified differently when at the same distance. (EX-1001, claim 1.)

Violates Board's Preliminary Construction: "a user" means "one or more users"
(DI, 15.)

Contradicts PO's Expert:

15 A. All right. "A user" means one or more. So
16 it's not limited to one user. It could be more than
17 one.
18 Q. Okay. It could be just one person speaking
19 to the device, right?
20 A. It could be, sure.
21 Q. And that person's voice is classified into
22 one of two categories, right?
23 A. Correct.

(EX-1201, 30:7-23.)

PO Argument 4 Fails

PO’s Argument 4: Shin does not disclose “classifying” a voice based on distance. (POR, 26-29.)

’337 Patent:

30 the mel-frequency cepstrum coefficient. In these cases, the voice classifying unit 38 sets a threshold of the feature value, and classifies the voice V1 as either one of the first voice V1A or the second voice V1B based on whether the feature value exceeds the threshold.

(EX-1001, 8:30-34.)

First Voice: $D \leq x$ (threshold)
Second Voice: $D > x$ (threshold)

Shin:

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
	$1\text{ m} \leq D < 2\text{ m}$	Content of first abbreviation level
	$2\text{ m} \leq D < 3\text{ m}$	Content of second abbreviation level
Second Voice	$3\text{ m} \leq D < 4\text{ m}$	Content of third abbreviation level
	$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

(EX-1003, Table 3 (annotated); EX-1002 ¶64.)

“The left column of Table 3 includes several categories (or classifications) into which the distance may fall. And the right column of Table 3 describes the result of each such classification on the device’s output. ... This is exactly how the ’337 patent describes voice classification.”

(EX-1202 ¶¶18-19.)

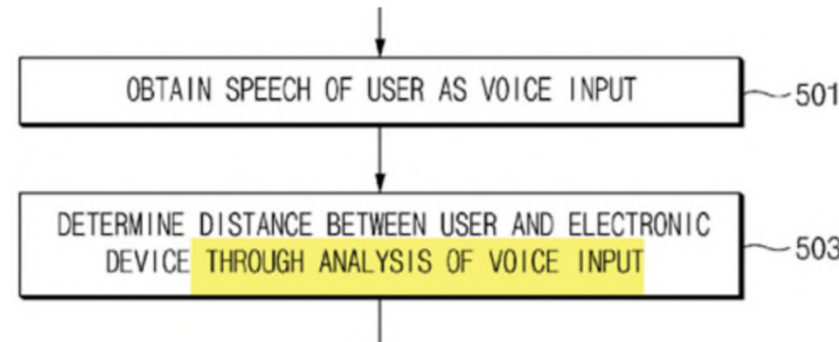
PO Argument 5

PO’s Argument 5: Shin does not disclose “analyzing” a voice to classify it based on distance. (POR, 29-31)

Contradicts ’337 Specification: the “analysis” is comparing the distance to a threshold. (EX-1001, 8:30-34; EX-1202 ¶15.) Shin discloses the same thing. (EX-1202 ¶21.)

Ignores Shin’s Express Disclosure:

1. “processor 120 may **analyze a voice** input ... and may output content corresponding to the voice input[.]” (EX-1003 ¶[0048].)
2. “processor 120 ... obtains a volume of the speech ... **by analyzing the voice input**” and may determine the user-to-device distance “based on the distance derived from the volume of the speech and the distance computed by the distance detection module” [proximity sensor] to improve accuracy. (EX-1003 ¶¶[0076]-[0077].)



(EX-1003, Fig. 5A (excerpt); EX-1202 ¶22.)

PO Argument 6

PO's Argument 6: Classification cannot be based on distance alone. (PO Sur-Reply, 13.)

and others. POR, 15; EX-1001, 7:64-8:34. By way of example, the specification describes classifying “the voice V1 to the second voice V1B when the voice V1 is *determined to be a whisper* as a result of analysis,” and “to the first voice V1A when the voice V1 is *determined not to be a whisper*.” EX-1001, 7:51-63(emphasis added); POR, 15-16 (emphasis changed). This is done by using combinations of “feature values” indicative of the characteristics of the voice. While the specification exemplifies “distance” as one such “feature values[]” of many, the plain language of the claims requires distance to be used in conjunction with other “features values” to classify voices.

PO Argument 6

Contradicts the Claim Language:

“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-1001, claim 1.)

PO Argument 6

Contradicts the Specification:

1. “the voice classifying unit 38 can perform the classification to the first voice V1A and the second voice V1B **by using any method.**” (EX-1001, 8:10-12.)
1. “distance can be used as **a feature value** to perform the classification to the first voice V1A and the second voice V1B” (EX-1001, 8:20-26.)
2. “the voice classifying unit 38 sets a threshold of **the feature value**, and classifies the voice V1 ... based on whether **the feature value** exceeds the threshold” (EX-1001, 8:30-34.)

PO Argument 6

Contradicts Both Experts:

Petitioner's Expert:

1. “a POSITA would have understood the specification to disclose that voice can be classified ‘by using any method,’ including distance alone.” (EX-1202 ¶16; *id.* ¶12.)

Patent Owner's Expert:

1. Some embodiments “use the ... distance between the speaker and the device” to “classify the voice”. (EX-2013 ¶38.)
2. Specification compares “user-to-device distance” and a threshold “for the purpose of distinguishing one class of voice from another.” (EX-2013 ¶52.)
3. Each characteristic described in the specification, including distance, “**can be used alternatively**” to “classify voices.” (EX-2013 ¶53.)
4. Distance is one of the “classes of separately identifiable voices”. (EX-1201, 64:6-19.)

PO Argument 6

Contradicts Patent Owner's Representations in District Court:

1. “the claim uses distance to classify the speaker” (EX-1203, 45:4-10.)
2. The device “operates using distance to classify voices” (EX-1203, 54:6-7.)
3. “Classifying the incoming voice as a first recorded voice or second recorded voice, that’s a binary decision ... [and] we did it with the distance calculation.” (EX-1203, 45:4-10.)

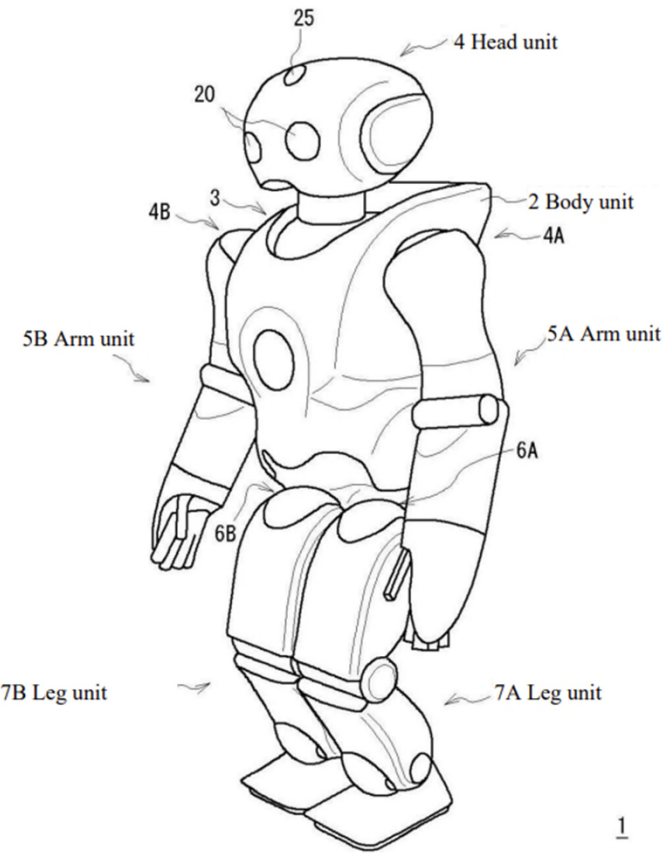
Contradicts Patent Owner's Complaint in District Court:

1. Asserted that a voice would satisfy the claim limitation if the voice is classified as “a near/close voice” (first voice) or a “far/distant voice” (second voice).
(EX-2016, 141:2-143:10; *SoundClear Techs. LLC v. Amazon.com, Inc.*, No. 1:24-cv-01283, Dkt. 1 at 14 (E.D.Va. July 25, 2024).)

Shimomura

Shimomura Overview

[FIG. 1]



(EX-1004, Fig. 1.)

[0050]

The robot 1 is equipped with a speech control function that controls speech forms such as speech volume, speech speed, intonation, and intervals between phrases according to the physical distance to the conversation partner. In

(EX-1004 ¶[0050].)

[0058]

In this case, the speech character string transformation unit 51 stores and holds in the ROM 33 (Fig. 3) a character string transformation determination table 50 that specifies whether or not to transform the character string to insert "ne" after each phrase according to the distance to the conversation partner, as shown in FIG. 7.

(EX-1004 ¶[0058].)

[0060]

When the speech character string transformation unit transformation unit 51 is given character string and distance information from the behavior selection control unit 45, it judges whether or not to transform the character string based on the character string information according to the character string transformation processing procedure RT2 shown in FIG. 8 based on the distance information and the character string transformation determination table 50, and transforms the character string if necessary.

(EX-1004 ¶[0060].)

Shimomura Discloses Classifying Voice Based on Distance

[0058]

In this case, the speech character string transformation unit 51 stores and holds in the ROM 33 (Fig. 3) a character string transformation determination table 50 that specifies whether or not to transform the character string to insert "ne" after each phrase according to the distance to the conversation partner, as shown in FIG. 7.

(EX-1004 ¶[0058].)

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

(EX-1004, Fig. 7.)

Shimomura Discloses Classifying Voice Based on Distance

[FIG. 7]

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

(EX-1004, Fig. 7 (annotated); EX-1002 ¶139.)

“Shimomura discloses classifying voice into two categories based on the user-to-device distance: (1) a first voice when the distance is greater than 350 cm, and (2) a second voice when the distance is less than 350 cm.”

(EX-1002 ¶139.)

Shimomura Discloses Generating the Claimed Output Sentences

[0062]

Specifically, the speech character string transformation unit 49 judges that the character string should not be transformed if the distance to the conversation partner recognized by the image recognition unit 41 is less than 350 [cm] according to the character string transformation determination table 50, and judges that the character string should be transformed if the distance to the conversation partner is 350 [cm] or more.

(EX-1004 ¶[0062].)

[0065]

The speech character string transformation unit 49 then proceeds to step SP13, where it sequentially adds the character "ne" to the end of each clause-final auxiliary verb detected in step SP12. Therefore, when the speech character string transformation unit 49 is given the character string "Today's weather in Tokyo is sunny" from the behavior selection control unit 45, it will generate 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].)

[0092]

"ne" is inserted may be changed in stages according to the distance to the conversation partner. It goes without saying that words other than "ne" can be used as words to be inserted after each phrase end to create a gap between phrases.

(EX-1004 ¶[0092].)

Shimomura Discloses Generating the Claimed Output Sentences

159. 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 ¶159.)

Shimomura Outputs the Sentences at Different Volumes

[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

(EX-1004, Fig. 9.)

[0070]

Specifically, when the distance to the conversation partner given as distance information is, **for example**, less than 50 cm, the speech synthesis unit 48 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. The speech synthesis unit 48 then changes the parameter values of the volume parameters as necessary according to the result of this determination.

(EX-1004 ¶[0070].)

Obvious Modifications to Shimomura

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

(EX-1004, Fig. 7.)

[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

(EX-1004, Fig. 9.)

Obvious Modification #1 to Shimomura

[FIG. 9]

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

FIG. 9 Volume Change Table

(EX-1004, Fig. 9.)

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

(EX-1002 ¶163.)

Obvious Modification #1 to Shimomura

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

Less than 350 cm	350 cm or more
None	Implementation

(EX-1004, Fig. 7 (annotated); EX-1002 ¶171.)

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

First Modified Volume Change Table

(EX-1002 ¶171.)

Obvious Modification #2 to Shimomura

[FIG. 9]

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

FIG. 9 Volume Change Table

52

(EX-1004, Fig. 9.)

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

(EX-1002 ¶163.)

Motivations to Modify Shimomura

1. Simple substitution of a known element (350 cm threshold as in Fig. 7 and 10-11) to obtain predictable result (device that adjusts output volume based on whether user is closer or farther than 350 cm).
2. Combination of known elements (distance thresholds) to achieve expected result.
3. Application of known technique (adjusting output volume at threshold) to achieve predictable result.
4. Routine experimentation and common sense to ensure audibility at 350 cm given Shimomura's goal of enhancing user interactions by adjusting volume with distance.
5. The threshold is indicated in Figures 7 and 10-11.

Motivations to Modify Shimomura

6. Obvious to try based on environment:

169. Sixth, such a modification would have been obvious to try. In environments where Shimomura's device operates at varied distances, users may interact from beyond 250 cm. 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. This approach simplifies design by minimizing programming complexity. 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. Shimomura's existing incremental volume changes make this a predictable solution for ensuring audibility at greater distances.

PO's Arguments Fail

PO Argument: Shimomura does not “classify” voice and does not “analyze” voice.
→ Fails for the same reasons as for Shin.

PO Argument: Shimomura teaches away from the proposed modification.
→ Shimomura does not explicitly discourage the modification.
→ Based on incorrect assumptions.

PO never rebuts Petitioners arguments.

PO concedes that a POSITA would have been motivated to change the thresholds:

distance. *Id.*, ¶104 (citing EX-1004, ¶¶[0072]-[0073]). A POSITA also would have understood that, in addition to changing speech volume, other speech forms, such as speech speed, pause length, phrase-end emphasis, and speech character string could also have been changed depending on user-to-device distance to make Shimomura’s robot’s speech easier to understand. *Id.* (citing EX-1004, ¶¶[0060]-
(POR, 48-49.)

Ground 4: Shimomura + Shin

1. Shin's volume change table (Table 1) and output content table (Table 3) include matching distance ranges, correlating specific distances with volume levels and information amounts.
2. Both references state that the tables/figures are provided as examples.
3. Shin discloses that information may be disclosed "dichotomously" and a POSITA would have been motivated to use this approach to simplify Shimomura's design.

Ground 5: Shin/Shimomura + Kristjansson

Neither PO nor its expert dispute that Kristjansson discloses the relevant limitations or that a POSITA would have been motivated to combine the teachings of Kristjansson with Shin and/or Shimomura.

(POR, 54-55; EX-2013 ¶¶114-15.)

Appendix

Claim 1

1. A voice-content control device, comprising:

- [a] a proximity sensor configured to calculate a distance between a user and the voice-content control device;
- [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;
- [c] a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user;
- [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
- [e] an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto, wherein the voice-content generating unit is further configured to
 - [f] generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice, and
 - [g] 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
- [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.

Claims 2-3

2. The voice-content control device according to claim 1, wherein

[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

[b] the voice-content generating unit is further configured to generate the text data including the acquisition content information as the output sentence.

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. A voice-content control method, comprising:

- [a] calculating a distance between a user and a voice-content control device;
- [b] acquiring a voice spoken by a user;
- [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;
- [d] analyzing the acquired voice to execute processing intended by the user;
- [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
- [f] adjusting a sound volume of voice data obtained by converting the output sentence thereinto, wherein
- [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
- [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. A non-transitory storage medium that stores a voice-content control program that causes a computer to execute:
- [a] calculating a distance between a user and a voice-content control device;
 - [b] acquiring a voice spoken by a user;
 - [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;
 - [d] analyzing the acquired voice to execute processing intended by the user;
 - [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
 - [f] adjusting a sound volume of voice data obtained by converting the output sentence thereinto, wherein
 - [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
 - [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.

Petitioners' Demonstratives

May 20, 2026

Amazon.com, Inc. v. SoundClear Technologies, LLC.

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

