

Petitioners' Demonstratives

July 8, 2026

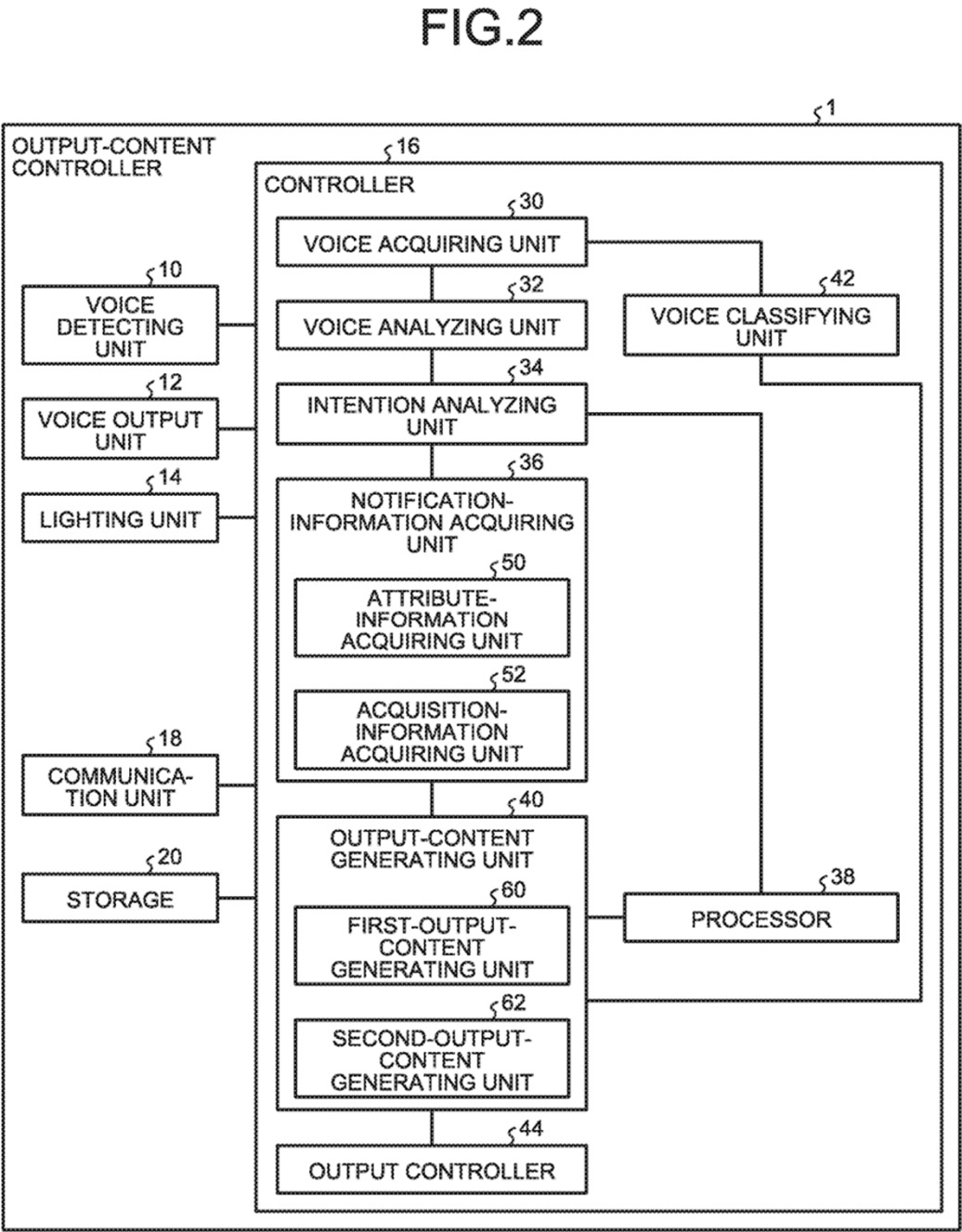
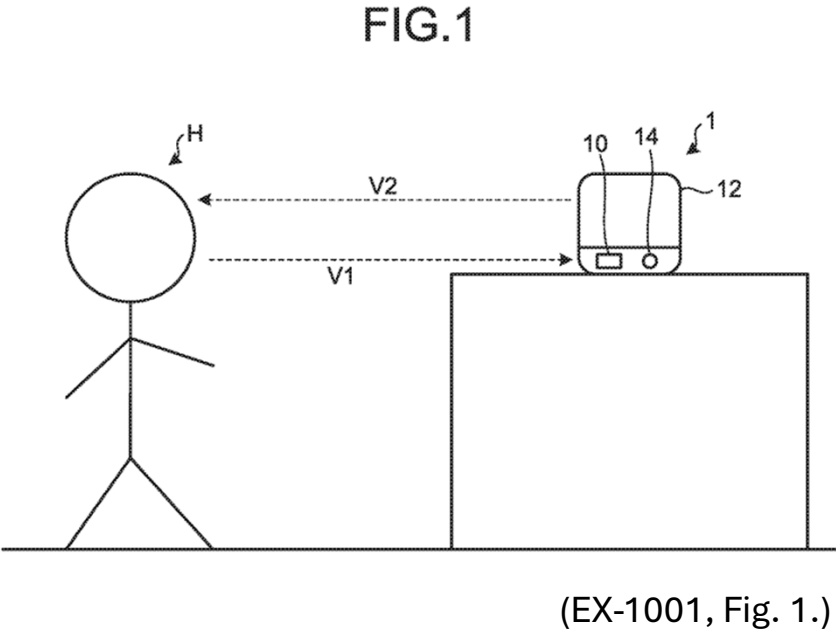
Amazon.com, Inc. v. SoundClear Technologies, LLC

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



<i>Amazon v. SoundClear</i> US Patent 11,244,675 Amazon EX-1301

'675 Patent Figures



(EX-1001, Fig. 2.)

Claim 1

1. An output-content control device comprising:
 - [a] **a voice acquiring unit** configured to acquire a voice spoken by a user;
 - [b] **a voice classifying unit** configured to calculate a distance between the user and the output-content control device by **a proximity sensor** to classify the voice into either a first voice or a second voice based on the calculated distance;
 - [c] **an intention analyzing unit** configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;
 - [d] **a notification-information acquiring unit** configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and
 - [e] **an output-content generating unit** configured to generate an output sentence as sentence data to be output to the user based on the notification information, wherein the output-content generating unit is further configured to
 - [f] *generate*, when the voice is determined to be the first voice, *a first output sentence* in which at least one word selected among words included in the content information of the notification information is replaced with another word, and to
 - [g] *generate*, when the voice is determined to be the second voice, *a second output sentence* which includes all of the intention information and the content information.

Claim 1

1. An output-content control device comprising:

[a] a **voice acquiring unit** configured to acquire a voice spoken by a user;

[b] a **voice classifying unit** configured to **calculate a distance** between the user and the output-content control device by a **proximity sensor** to **classify the voice into either a first voice or a second voice based on the calculated distance**;

[c] an **intention analyzing unit** configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;

[d] a **notification-information acquiring unit** configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and

[e] an **output-content generating unit** configured to generate an output sentence as sentence data to be output to the user based on the notification information, wherein the output-content generating unit is further configured to

[f] *generate*, when the voice is determined to be the first voice, **a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word**, and to

[g] *generate*, when the voice is determined to be the second voice, *a second output sentence* which includes all of the intention information and the content information.

Claim 1: Key Limitations

1. Classify voice based on distance
2. Output:
 - First Voice → replace word(s)
 - Second Voice → full information

Grounds

Ground	Claims	Basis	References
1a	1-7	§ 103	Shin + Aoyama
1b	1-2, 5-7	§ 103	Shin + Iwase
2a	1-7	§ 103	Shimomura + Aoyama
2b	1-2, 5-7	§ 103	Shimomura + Iwase
3	3-4	§ 103	Shin or Shimomura + Iwase + Aoyama
4	5	§ 103	Grounds 1-2 + Schuster
5	1-7	§ 103	Grounds 1-4 + Kristjansson

Classify Voice Based on Distance

Relevant Disclosure in '675 Patent

the threshold. Note that the voice classifying unit 42 can perform the classification to the first voice V1A and the second voice V1B by using any method. For example, the
25 voice classifying unit 42 can perform the classification to the first voice V1A and the second voice V1B by using a slope of the peak in the spectrum as a feature value. Moreover, the voice classifying unit 42 can perform the classification to the first voice V1A and the second voice V1B by using either
30 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 a person and a wind noise as a feature value. Furthermore, a proximity sensor can be provided in the output-content control device 1, a distance between the user H and the
35 output-content control device 1 is calculated from a detection result of the proximity sensor, and the distance can be used as a feature value to perform the classification to the first voice V1A and the second voice V1B. Moreover, the voice classifying unit 42 can derive a Mel-frequency cepstrum coefficient as a feature value, and can perform the
40 classification to the first voice V1A and the second voice V1B based on the mel-frequency cepstrum coefficient. In these cases, the voice classifying unit 42 sets a threshold of the feature value, and classifies the voice V1 to either the
45 first voice V1A or the second voice V1B based on whether the feature value exceeds the threshold.

(EX-1001, 10:22-46.)

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



Shin Overview

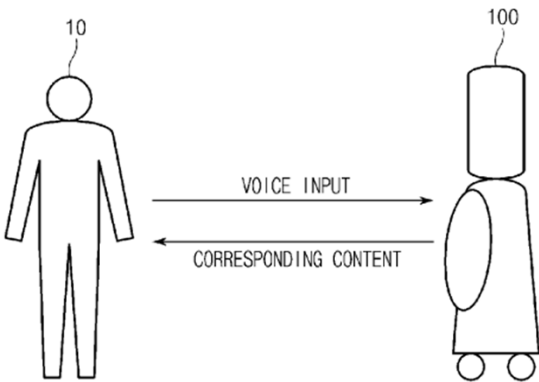


FIG. 1A
(EX-1003, Fig. 1A.)

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

(EX-1003, Table 3 (annotated); EX-1002 ¶61.)

“Shin discloses classifying the voice as a ‘first voice’ when the distance is between 1 to 2 meters, and a ‘second voice’ when the distance is less than 1 meter[.]”

(EX-1002 ¶61.)

Shin Discloses Classifying the Voice Based on Distance

'675 Patent:

V1B based on the mel-frequency cepstrum coefficient. In these cases, the voice classifying unit 42 sets a threshold of the feature value, and classifies the voice V1 to either the first voice V1A or the second voice V1B based on whether the feature value exceeds the threshold.

(EX-1001, 10:42-46.)

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

Shin:

TABLE 3

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

(EX-1003, Table 3 (annotated); EX-1002 ¶61.)

“The left column of Table 3 includes several categories (or classifications) into which the distance D 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 '675 patent describes voice classification.”

(EX-1202 ¶¶12-13.)

PO's Primary Argument

Voice classification cannot be based solely on distance
because “distance thresholds do not alone delineate one
class of voice from another”

PO's Argument Contradicts the Claim Language

“***classify the voice*** into either a first voice or a second voice ***based on the calculated distance***”

(EX-1001, claim 1.)

PO's Argument Contradicts the Specification

“*distance* can be used as *a feature value to perform the classification* to the first voice V1A and the second voice V1B”

(EX-1001, 10:32-38.)

“the voice classifying unit 38 sets a threshold of *the feature value*, and *classifies the voice ... based on whether the feature value exceeds the threshold*”

(EX-1001, 10:42-46.)

PO's Argument Contradicts PO's Complaint

63. On information and belief, one or more components of the Amazon Products and Services is, employs, or provides a voice-content control method comprising analyzing the acquired voice (*e.g.*, the signal representing the voice sound) to classify the acquired voice as either one of a first voice (*e.g.*, the voice of a particular first person, or a near/close voice) and a second voice (*e.g.*, the voice of a particular second person, or a far/distant voice) based on the distance between the user and the voice-content control device. On information and belief and as

* * *

calendar, etc. when the user's voice is detected. *See id.* Further, on information and belief and as another example, one or more components of the Amazon Products and Services classify a voice signal as a first voice (near) or a second voice (far) using, *e.g.*, the adaptive content feature or the SIR beam selection feature. *See*

(EX-1212 ¶63.)

PO's Argument Contradicts PO's Complaint

75. On information and belief, one or more components of the Amazon Products and Services is, employs, or provides an output-content control method comprising calculating a distance (*e.g.*, using ultrasound motion detection, adaptive content, and/or an algorithm for detecting the location of the voice audio source (*e.g.*, “SIR beam selection”)) between the user (*e.g.*, a person within range of an Echo device) and an output-content control device (*e.g.*, an Echo device) by a proximity sensor (*e.g.*, hardware/software associated with ultrasound motion detection, adaptive content, and/or an algorithm for detecting the location of voice audio sources) to classify the voice (*e.g.*, the signal representing the voice sound) into either a first voice (*e.g.*, the voice of a particular first person, or a near/close voice) or a second voice (*e.g.*, the voice of a particular second person, or a far/distant voice) based on the calculated distance. On information and belief and as an example, one or more of the Amazon Products and Services use “adaptive content” to “detect the proximity of a person.” *See* <https://www.amazon.com/gp/help/customer/display.html?nodeId=TBmEUIW1U7xwcgPvdl>. On

(EX-1212 ¶75.)

PO's Argument Contradicts PO's Complaint

Further, on information and belief and as an example, Amazon Products and Services can display more or less information depending on whether an acquired voice has been classified as one that corresponds to a nearby/close user or a far/distant user (e.g., using Adaptive Content). On information and belief, the Amazon Products and Services can also provide a voice response with more or less information consistent with the more or less information that is visually displayed. *See*

(EX-1212 ¶78.)

PO's Argument Contradicts PO's Prior Statements

“*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.)

PO's Argument Contradicts PO's Expert

The “feature values’ (including user-to-device distance) and threshold thereof are determined and set for the purpose of distinguishing one class of voice from another.”

(EX-2016 ¶45.)

The “voice classifying unit sets certain feature values,” such as distance, “that distinguish between different voices”

(EX-2016 ¶40.)

PO's Argument Contradicts Petitioners' Expert

“[T]he voice classifying unit may classify the user’s voice based ***solely on the user-to-device distance.***”

(EX-1202 ¶15; see also EX-1002 ¶59.)

“The specification ... explicitly discloses classifying ***based on distance alone.***”

(EX-1202 ¶16.)

PO's Secondary Argument

The prior art references “classify distances, not voices”

Distance Is a Voice Characteristic

The prior art discloses classes or categories of distances ***from which the voice is received.***

PO's Expert Admits Distance Is a Voice Characteristic

“distance” is an “aspect” or “characteristic” of human voice

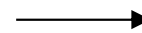
(EX-2016 ¶137.)

PO Admitted Distance Is a Voice Characteristic

IPR2025-00673 ('675 Patent)

“distance” is a “feature value” but “has no direct relationship with any voice attributes/characteristics.”

(PO Sur-Reply, 3.)



IPR2025-00565 ('337 Patent)

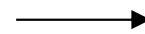
“the specification exemplifies ‘distance’ as” a “feature value” that is “indicative of the characteristics of the voice”

(‘337 POR Sur-Reply, 13.)

PO Admitted Distance Is a Voice Characteristic

IPR2025-00673 ('675 Patent)

“Distance ***is not*** an aspect of a voice.”
(POR, 7.)



IPR2025-00565 ('337 Patent)

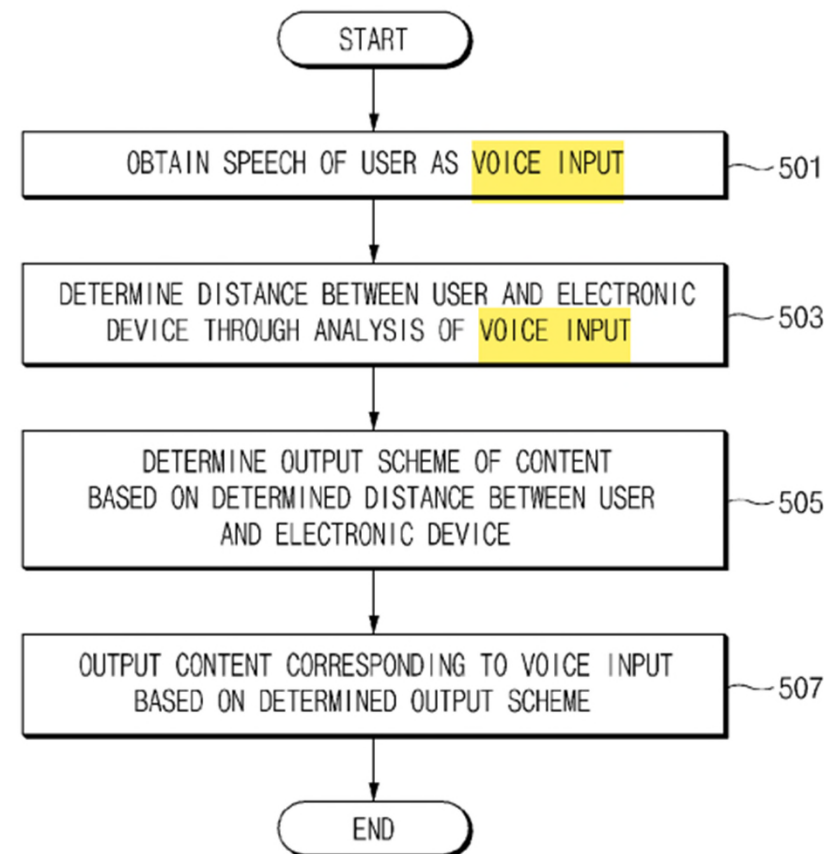
“[D]istance ***is*** ... one aspect of voice.”
(’337 POR, 14.)

POR Argument 1 Fails

POR Argument 1: Shin only mentions “voice” in connection with the output. (POR, 23-24.)

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 ¶11.)

POR Argument 2 Fails

POR 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.)

POR Argument 3 Fails

POR Argument 3: Shin's device would not distinguish between the voices of two different users standing the same distance from the device. (POR, 25-26.)

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.)

Contradicts PO's Expert:

19	Q.	You agree that the claim only requires
20		one user to use the device, correct?
21	A.	Generally speaking, yes, it does
22		require at least one user. The fact that it's
23		spoken by a user, that could be one or more users.

(EX-1201, 13:19-14:2.)

Word Replacement

'675 Patent Example

User: What is Mr. Yamada's schedule today?

Second output sentence (second voice):

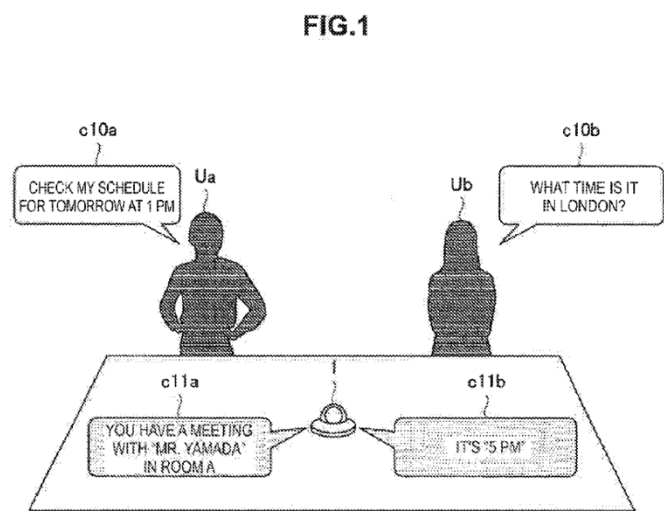
“Mr. Yamada's schedule on Mar. 20, 2020 is a meeting with Mr. Yoshida from 15 o'clock at Tokyo building.”

First output sentence (first voice):

“Mr. Yamada's schedule on Mar. 20, 2020 is a meeting with ~~Mr. Yoshida~~ **Yoshi** from 15 o'clock at Tokyo building.”

(EX-1001, 14:23-43.)

Aoyama Overview



(EX-1010, Fig. 1.)

FIG.14

The diagram shows a user interface for a user with ID 000001. It displays a table of personal name data. The table has two columns: FORMAL NAME and NAME. The rows are Michael Smith (Mickey), Elizabeth Green (Betty), and Katharine McPhee (Kate). The table is labeled d10 and d11.

FORMAL NAME	NAME
Michael Smith	Mickey
Elizabeth Green	Betty
Katharine McPhee	Kate

(EX-1010, Fig. 14.)

Aoyama's Disclosure

User: “Check my schedule for tomorrow at 1 pm.”

Second output sentence (second voice):

“You have a meeting with Taro Yamada in room A at 13:00 10/1/2014.”

First output sentence (first voice):

“You have a meeting with ~~Taro~~ **Mr.** Yamada in room A ~~at 13:00 10/1/2014~~ **tomorrow at 1 pm.**”

(EX-1010 ¶¶[0138]-[0140].)

PO admits Aoyama discloses word replacement. (POR, 30-31.)

A POSITA Would Have Combined Shin and Aoyama

Many Motivations to Combine (EX-1002 ¶¶97-102):

1. The combination would further Shin's primary goal of simplifying output for users at greater distances and making the information easier for users to grasp (e.g., replacing full names with simple nicknames or complicated date/time formats with simpler ones).
2. Aoyama teaches replacing words based on user state to make user-device interaction "more natural," which would have motivated a POSITA to use a proximity sensor and replace words based on distance.
3. "Given known privacy concerns in voice systems like Shin's, a person of ordinary skill would have been motivated to incorporate Aoyama's word replacement feature into Shin's device to replace words containing personal information with neutral alternatives."
4. The combination "merely represents the addition of one known element (Aoyama's word replacement) to another known element (Shin's distance-based content abbreviation/simplification) to obtain predictable results (providing content to users at greater distances that is abbreviated or simplified using word replacement)."
5. Shin's device presents only two design options for abbreviating or simplifying output for users at greater distances: omitting words or replacing them. Thus, Aoyama's word replacement approach would have been obvious to try.

PO's Arguments Fail

PO's Argument: Aoyama teaches away from word replacement based on distance because its word replacement depends only on the user's "speaking style."

(POR, 31-32.)

But:

- (a) Aoyama is not so limited
- (b) Aoyama does not teach away – nothing in Aoyama discourages replacing words based on distance

(EX-1202 ¶17.)

PO's Argument: Aoyama's teaching of adjusting "amount of information output" is similar to Shin's and thus a POSITA would have adjusted the amount of information rather than replacing words. (POR, 32-33.)

But:

- (a) Legally incorrect—prior art must be considered for everything it teaches
- (b) Aoyama teaches that the system "may achieve other effects that are clear to those skilled in the art"

(EX-1010 ¶[0256].)

PO's Arguments Fail

PO's Argument: Word replacement based on distance “would have led to a response with which ‘the user may feel uncomfortable.’” (POR, 32.)

But:

PO relies on a single example—changing response from standard time to military time based on distance. Aoyama discloses many other examples:

1. Referring to users by different names (Papa and Mama instead of given names)
2. Replacing “Taro Yamada” with “Mr. Yamada”
3. Replacing dates of “9/30/2014” with an expression such as “today,” “tomorrow,” or “the day after tomorrow.”

(EX-1010 ¶¶[0221], [0138]-[0140].)

Iwase

Iwase's Disclosure

User: Asks about the weather.

Second output sentence (second voice):

“I recommend bringing an umbrella.”

First output sentence (first voice):

“~~I recommend bringing~~ You should take an umbrella.”

(EX-1017 15:6-14.)

PO admits Iwase discloses word replacement. (POR, 39-40.)

A POSITA Would Have Combined Shin/Shimomura and Iwase

Many Motivations to Combine (EX-1002 ¶¶145-150):

1. Iwase suggests the combination because it teaches that its word replacement system allows a user “to grasp notification contents more effectively.”
(EX-1017 2:11-14, 15:24-28, 21:66-22:8.)
2. Iwase’s configuration would have been obvious to try because Shin’s device presents only two options for modifying output: removing or replacing words.
3. The references are similar in that they each disclose similar outputs.
(EX-1003 ¶[0088]; EX-1017, 15:6-14.)
4. The references are similar in that they each disclose determining user-to-device distance and altering output based on that distance.
(EX-1017 9:59-10:9, 11:32-39, 12:4-9.)
5. The configuration “represents the simple addition of one known element (Iwase’s word replacement) to another known element (Shin’s distance-based output adjustment) to obtain predictable results (easier-to-understand output to users at greater distances).”

PO's Arguments Fail

PO's Argument: A POSITA would not have been motivated to combine Shin or Shimomura with Iwase because “Iwase is not concerned at all with distance.” (POR, 41.)

But:

In addition, for example, in a case where the factor is a distance to the user, a value (cm, etc.) of the distance recognized by the image analysis unit 320, or a level value (for example, 1 to 10, etc.) corresponding to the distance, 35 may be input into the detection value F_i . In other words, the situation estimation unit 330 estimates that the further the user is away from the information processing terminal 10, it is more difficult to acquire user's attention.

(EX-1017 11:32-39.)

PO's Argument: A POSITA would not have been motivated to modify Shin in view of Iwase because “abbreviated information sufficed for Shin's purposes.” (POR, 41.)

But:

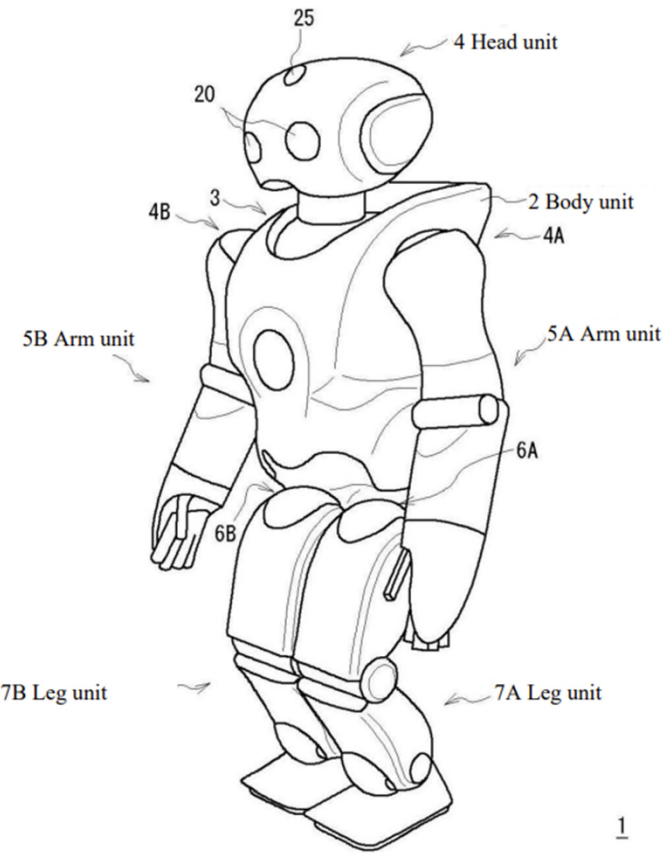
“[J]ust because [Shin and Iwase] met the same need ... in different ways would not mean that a POSA would reject the idea of modifying [Shin's] way of doing so by adding [Iwase's] solution.”

Stryker Eur. Operations Holdings LLC v. OsteoMed LLC, No. 23-2397, 2025 WL 2249005 at *8 (Fed. Cir. Aug. 7, 2025).

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 ¶170.)

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

(EX-1002 ¶170.)

Claim 5: Whisper

Claim 5: Claim Construction

Claim 5: The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice **when the voice spoken by the user is a whisper.**

Petitioners: Claim 5 requires that the device be capable of classifying the voice in circumstances where the user whispers to the device. (See Pet., 37-40, 58-60, 65-67; Reply, 16-18.)

PO: Claim 5 requires the device to determine that the user's voice is a whisper.
(POR, 33-36; PO Sur-Reply, 4-5.)

The Board does not need to construe the limitation because the prior art discloses it under either construction.

Claim 5: Prior Art

Shin and Shimomura's Devices Process Whispered Speech: A POSITA would have understood Shin and Shimomura's devices to process whispered speech, or would have been motivated to modify them to do so.

(Pet., 37-40, 58-60; EX-1002 ¶¶132-37, 198-202.)

Schuster Discloses Detecting Whispered Speech:

For example, in an embodiment, a mobile communication
60 device enters a private communication mode with the user
when the user's preceding voice interaction with the device
is whispered. The private communication mode ("private
mode") allows the user to interact with the device in a
private manner, to avoid disturbing other people, to keep
65 their interaction confidential, or simply to avoid social
awkwardness associated with communicating with an elec-
tronic device.

To detect the Intended Private Interaction, the device may
use one or a combination of several methods. In an embodi-
ment, the signal received at the microphone is analyzed to
detect a whispered microphone input. As used herein, the
terms "whisper" and "whispered" refer to unvoiced or 5
substantially unvoiced speech as well as hushed speech,
which may be spectrally distinct from either voiced speech
or whispered speech. Continuing, any suitable method may
be used in this embodiment to detect whispered input. For
example, cepstral analysis applied to the incoming micro- 10
phone signal may be used to detect whispered input.

(EX-1016, 2:59-3:11.)

Claim 5: Motivation to Combine

Simple Addition of Known Elements: Configuring Shin/Shimomura's system to classify a voice based on whisper is merely the simple addition of a feature from Schuster (detecting whisper) to Shin/Shimomura's voice classification systems and predictably results in a system that classifies voice based on whisper and/or distance.

(Pet., 66-67; EX-1002 ¶1229.)

Similarity of the References: Schuster discloses classifying voice based on distance, just like Shin and Shimomura. (Pet., 67; EX-1002 ¶1229.)

In addition to whispered audio input, camera image analysis can be used additionally or alternatively to detect 20 that the user intends to enter the private mode. For example, in an embodiment, image analysis is applied to captured images or video to detect a fixed gaze or other indicator of private intent. For example, in an embodiment, the camera is used to capture lip movement which is then used by the 25 device processor to interpret the movements into words. Alternatively or additionally, other sensors may be used as well to detect the user's desire to enter the private mode; for example, proximity sensors can be used to sense proximity of the user. 30

(EX-1016, 3:19-31.)

Grounds 3 and 5

Ground 3: Neither PO nor its expert dispute that Aoyama discloses the relevant limitations of claims 3 and 4 or that a POSITA would have been motivated to combine the teachings of Aoyama with Shin or Shimomura and Iwase.

(POR, 56; EX-2016 ¶154.)

Ground 5: 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, Shimomura, Aoyama, Iwase, and/or Schuster.

(POR, 59-60; EX-2016 ¶¶159-63.)

Appendix

Claim 1

1. An output-content control device comprising:

- [a] a voice acquiring unit configured to acquire a voice spoken by a user;
- [b] a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;
- [c] an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;
- [d] a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and
- [e] an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information, wherein the output-content generating unit is further configured to
 - [f] generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word, and to
 - [g] generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.

Claims 2-5

2. The output-content control device according to claim 1, wherein
 - [a] the notification information includes type information indicating a type of the content information, and
 - [b] the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be placed with the other word based on a word included in the content information and the type information that is associated with the content information.
3. The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.
4. The output-content control device according to claim 3, wherein
 - [a] the relationship information further includes information of type information indicating a type of the first word, and
 - [b] the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word.
5. The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.

Claim 6

6. An output-content control method, comprising:

- [a] acquiring a voice spoken by a user;
- [b] calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;
- [c] analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;
- [d] acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and
- [e] generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and
- [f] generating, when the voice is not determined to be the second voice, a second output sentence which includes all of the intention information and the content information.

Claim 7

7. A non-transitory storage medium that stores an output-content control program that causes a computer to execute:

- [a] acquiring a voice spoken by a user;
- [b] calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance;
- [c] analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user;
- [d] acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and
- [e] generating, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word; and
- [f] generating, when the voice is not determined to be the second voice, a second output sentence which includes all of the intention information and the content information.

Petitioners' Demonstratives

July 8, 2026

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

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

