

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

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

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner

Case No. IPR2025-00673 (Patent No. 11,244,675)

Patent Owner's Demonstratives

Introduction and Background

Grounds

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

Pet., 5

Challenged Patent



US011244675B2

(12) **United States Patent**
Naganuma

(10) **Patent No.:** US 11,244,675 B2
(45) **Date of Patent:** Feb. 8, 2022

(54) **WORD REPLACEMENT IN OUTPUT
GENERATION FOR DETECTED INTENT BY
VOICE CLASSIFICATION**

G06F 40/284; G10L 13/033; G10L
15/1822; G10L 15/22; G10L 25/63; G10L
2015/226; G10L 2015/228; G10L 17/00;
G10L 17/06

(71) Applicant: **JVC KENWOOD Corporation**,
Yokohama (JP)

USPC 704/1, 9, 10, 231, 246, 239, 270, 275,
704/238
See application file for complete search history.

(72) Inventor: **Tatsumi Naganuma**, Yokohama (JP)

(56) **References Cited**

(73) Assignee: **JVC KENWOOD Corporation**,
Yokohama (JP)

U.S. PATENT DOCUMENTS

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 152 days.

5,852,804 A 12/1998 Sako
10,341,825 B2* 7/2019 Dowlatkhah G10L 13/08
(Continued)

FOREIGN PATENT DOCUMENTS

(21) Appl. No.: **16/295,034**

JP 04-204700 7/1992
WO 2016/136062 9/2016

(22) Filed: **Mar. 7, 2019**

Primary Examiner — Martin Lerner

(65) **Prior Publication Data**

US 2019/0279631 A1 Sep. 12, 2019

(30) **Foreign Application Priority Data**

Mar. 12, 2018 (JP) JP2018-044598

(51) **Int. Cl.**

G10L 17/06 (2013.01)
G06F 40/247 (2020.01)
G10L 15/22 (2006.01)
G10L 15/18 (2013.01)
G06F 3/16 (2006.01)
G10L 25/51 (2013.01)

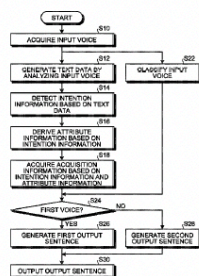
(52) **U.S. Cl.**

CPC **G10L 15/22** (2013.01); **G06F 3/16**
(2013.01); **G10L 15/1815** (2013.01); **G10L**
25/51 (2013.01); **G10L 2013/223** (2013.01)

(58) **Field of Classification Search**

CPC G06F 40/237; G06F 40/247; G06F 40/279;

7 Claims, 6 Drawing Sheets



-1-

Amazon v. SoundClear
US Patent 11,244,675
Amazon Ex. 1001

(54) WORD REPLACEMENT IN OUTPUT GENERATION FOR DETECTED INTENT BY VOICE CLASSIFICATION

(57) **ABSTRACT**

An output-content control device includes a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to determine whether the voice is a predetermined voice; 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

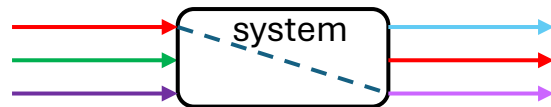
Ex. 1001, 1

Classification

Black Box



A system that receives [input] and [process] to [create result]



22 Would a POSITA have been able to determine
1 how a device classifies inputs simply by looking at
2 how the device adjusts outputs?

3 MR. REEVES: Objection. Form. Incomplete
4 hypothetical. Outside the scope.

5 THE WITNESS: I think, in general, the
6 answer would be no. I don't believe that that's
7 relevant to the claims of the present patent or the
8 '337 patent.

18 Wednesday, April 8, 2026

19 12:07 p.m.

20 Job No.: 627236

21 Pages 1 - 147

22 Reported By: Joan V. Cain

Ex. 2021, 62:22-63:8; POSR, 12-13

Transcript of Richard Stern, Ph.D.
Conducted on April 8, 2026

Dr. Richard Stern (for Petitioners)



Prior Art Classifies Distance

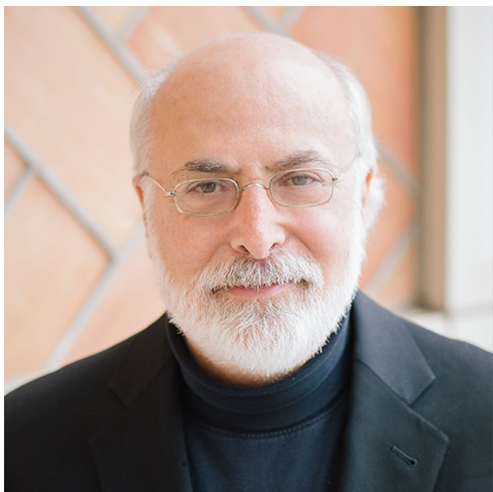
8 SOUNDCLEAR TECHNOLOGIES LLC,
9 Patent Owner.
10 IPR2025-00565
11 U.S. Patent No. 11,069,337
12 IPR2025-00673
13 U.S. Patent No. 11,244,675
14
15 VIDEOTAPED DEPOSITION OF
16 RICHARD STERN, PH.D.
17 Conducted Remotely Over Zoom

REMOTE VIDEOTAPED DEPOSITION OF
RICHARD STERN, PH.D.
December 10, 2025

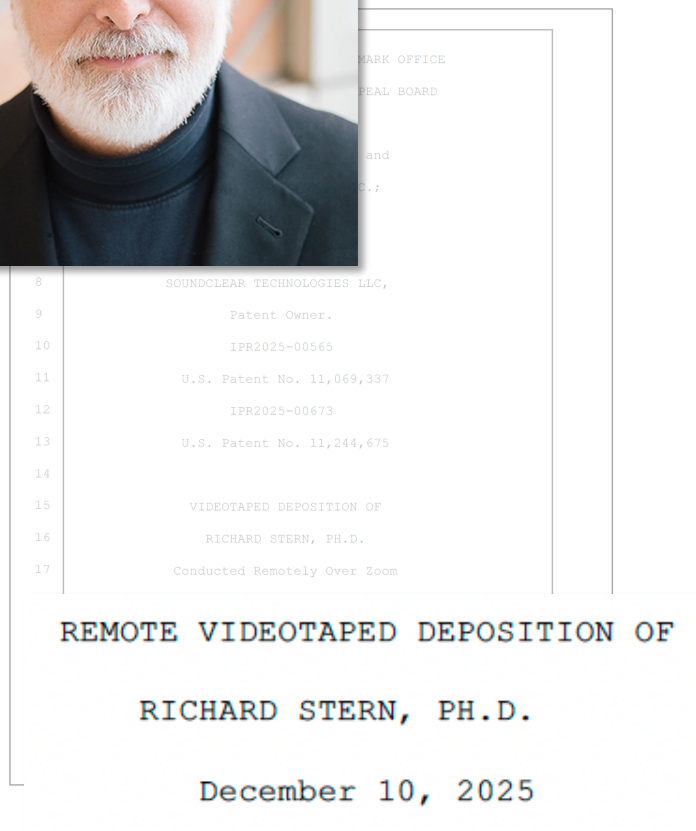
3 Q Okay. Now, referring to Table 3 of Shin,
4 as it appears in paragraph 61 of your declaration,
5 Table 3 classifies the distances between a user
6 and the electronic device, correct?
7 A Yes.

Ex. 2017, 48:3-7; POSR, 14-15

Dr. Richard Stern (for Petitioners)



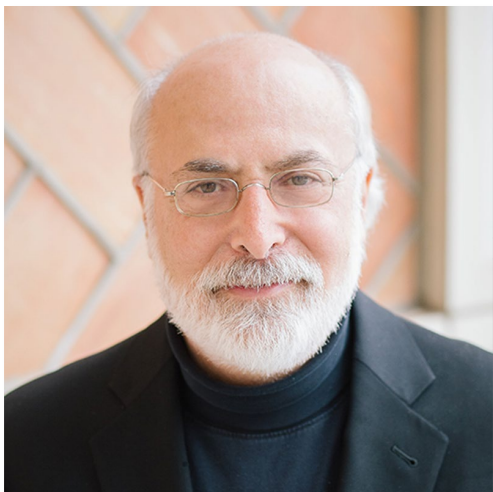
Prior Art Classifies Distance



7 Q Okay. Table 3 does not mention voices at
8 all, correct?
9 MR. REEVES: Objection, form.
10 A The table itself indeed only talks about
11 the distance. The independent variable is the
12 distance between the user and the device, but I
13 would have to review the specification to opine on
14 the context, but it is correct that Table 3 does
15 not refer to voices explicitly.

Ex. 2017, 49:7-15; POSR, 14-15

Dr. Richard Stern (for Petitioners)



Prior Art Classifies Distance

8 SOUND CLEAR TECHNOLOGIES LLC,
9 Patent Owner.
10 IPR2025-00565
11 U.S. Patent No. 11,069,337
12 IPR2025-00673
13 U.S. Patent No. 11,244,675
14
15 VIDEOTAPED DEPOSITION OF
16 RICHARD STERN, PH.D.
17 Conducted Remotely Over Zoom

REMOTE VIDEOTAPED DEPOSITION OF
RICHARD STERN, PH.D.
December 10, 2025

13 Q In the context of Shimomura, is there a
14 difference between classifying a voice versus
15 classifying a distance?
16 MR. REEVES: Objection, form.
17 A Well, it is hard for me to answer that,
18 because we never talk about classifying a voice.
19 We talk about classifying based on distance. The
20 input is always the physical measurement. The
21 output is the presentation of voice, but the
22 classification is based on the input.

Ex. 2017, 119:13-22; POSR, 10-11

Purpose and Target of Classification

“Classify the Voice into Either a First Voice or a Second Voice”

REMOTE VIDEOTAPED DEPOSITION OF

RICHARD STERN, PH.D.

December 10, 2025

7 Q Okay. So in this combination of Shin with
8 Iwase, do you -- is it your position that Iwase
9 classifies voices based on -- based on distance?

10 MR. REEVES: Objection, form, outside the
11 scope.

12 A As I noted when we reviewed all of these
13 references very briefly at the beginning of this
14 discussion, that all of the prior art citations
15 involved a certain amount of measuring what was
16 going on in the environment and using that to
17 modify content.

1. An output-content control device comprising:
a voice acquiring unit configured to acquire a voice
spoken by a user;
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;

Ex. 1001, 18 (cl. 1)

Ex. 2017, 95:7-17; POSR, 10-11

“Determine ... When the Voice Spoken by the User Is a Whisper”

The Specification

describes classifying input voice based on determination that the voice is a whisper

The voice classifying unit 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and determines whether the voice is a predetermined voice. In the present embodiment, the voice classifying unit 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and classifies the voice V1 to either one of a first voice V1A and a second voice V1B. The first voice V1A corresponds to a predetermined voice. In the present embodiment, the voice classifying unit 42 classifies the voice V1 to the first voice V1A when the voice V1 is determined to be a whisper as a result of analysis (determines as the predetermined voice).

Ex. 1001, 9:58-10:1; POR, 7-8; POSR, 4-5

“Classify the Voice into Either a First Voice or a Second Voice”

Classify “**a voice spoken by a user**” to delineate/distinguish between two classes of voices:

- “**a first voice**” or
- “**a second voice**”

POR, 5; POSR, 1-2

Ex. 1001, 18-19 (cls. 1, 6, and 7)

1. An output-content control device comprising:
a voice acquiring unit configured to **acquire a voice spoken by a user;**
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;

6. An output-content control method comprising:
acquiring a voice spoken by a user;
calculating a distance between the user and an output-content control device by a proximity sensor **to classify the voice into either a first voice or a second voice** based on the calculated distance;

7. A non-transitory storage medium that stores an output-content control program that causes a computer to execute:
acquiring a voice spoken by a user;
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;

“Classify the Voice into Either a First Voice or a Second Voice”

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**.

Ex. 1001, 19 (cl.5); POR, 33-36

1. An output-content control device comprising:
a voice acquiring unit configured to **acquire a voice spoken by a user**;
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;

Ex. 1001, 18 (cl. 1)

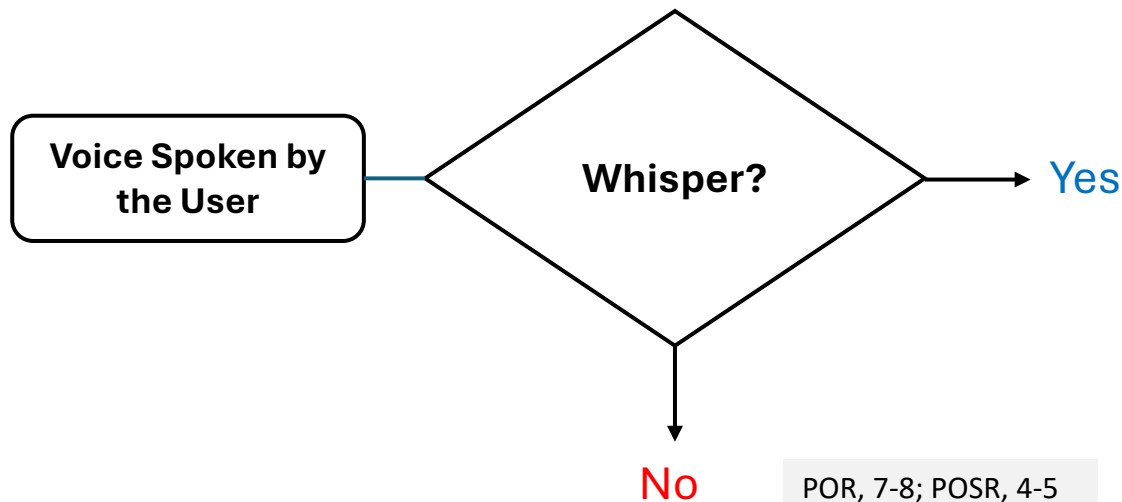
“Classify the Voice into Either a First Voice or a Second Voice”

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**.

Ex. 1001, 19 (cl.5); POR, 33-36

15 Q Okay. So -- so your position is that
16 claim 5 does not require the determination that a
17 voice is a whisper; is that correct?
18 A That's correct.

EX-2017, 93:15-18; POR, 34



PO argues that claim 5 requires determining that the voice is a whisper. (POR, 36.) It does not. Instead, claim 5 requires that the device be configured to determine that “the voice” is “the predetermined voice” in circumstances where the user whispers to the device. (EX-1001, claim 5; EX-1202 ¶19.) **Nothing in the claim requires the device to *determine* whether the received voice is a whisper.**

Reply, 16-17

“Feature Values” for Distinguishing One Class of Voice from Another

The voice classifying unit 42 classifies the voice V1 to the second voice V1B when the voice V1 is determined not to be a whisper. A whisper is an unvoiced voice spoken without vibrating the vocal chords, but it is not limited to the unvoiced voice spoken without vibrating the vocal chords as long as being classified to the first voice V1A by, for example, a method described later.

Furthermore, a proximity sensor can be provided in the output-content control device 1, a distance between the user H and the 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.

POR, 5-6; Ex. 1001, 10:1-8, 10:32-38; 10:47-53

Dr. Richard Stern (for Petitioners)



MARK OFFICE
FEAL BOARD
and
E.;

8 SOUNDCLEAR TECHNOLOGIES LLC,
9 Patent Owner.
10 IPR2025-00565
11 U.S. Patent No. 11,069,337
12 IPR2025-00673
13 U.S. Patent No. 11,244,675
14
15 VIDEOTAPED DEPOSITION OF
16 RICHARD STERN, PH.D.
17 Conducted Remotely Over Zoom
18 Wednesday, April 8, 2026
19 12:07 p.m.
20 Job No.: 627236
21 Pages 1 - 147
22 Reported By: Joan V. Cain

Transcript of Richard Stern, Ph.D.
Conducted on April 8, 2026

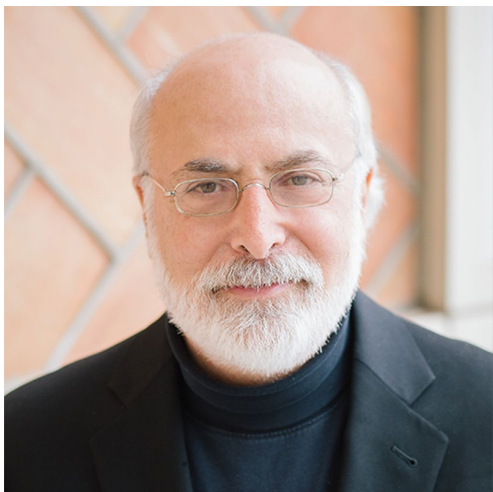
“Voice” is Distinct from “Distance”

- Distance has no direct relationship with any voice characteristics

1 Q Does the distance between the user and the
2 device alone tell you anything about the
3 characteristics of the user's voice?
4 MR. REEVES: Objection. Form. Outside
5 the scope.
6 A In general, no.

Ex. 2020, 54:1-6; POSR, 2-4; POR, 6

Dr. Richard Stern (for Petitioners)



MARK OFFICE
FEAL BOARD
and
E.;

8 SOUNDCLEAR TECHNOLOGIES LLC,
9 Patent Owner.
10 IPR2025-00565
11 U.S. Patent No. 11,069,337
12 IPR2025-00673
13 U.S. Patent No. 11,244,675
14
15 VIDEOTAPED DEPOSITION OF
16 RICHARD STERN, PH.D.
17 Conducted Remotely Over Zoom
18 Wednesday, April 8, 2026
19 12:07 p.m.
20 Job No.: 627236
21 Pages 1 - 147
22 Reported By: Joan V. Cain

Transcript of Richard Stern, Ph.D.
Conducted on April 8, 2026

“Voice” is Distinct from “Distance”

- Distance has no direct relationship with any voice characteristics

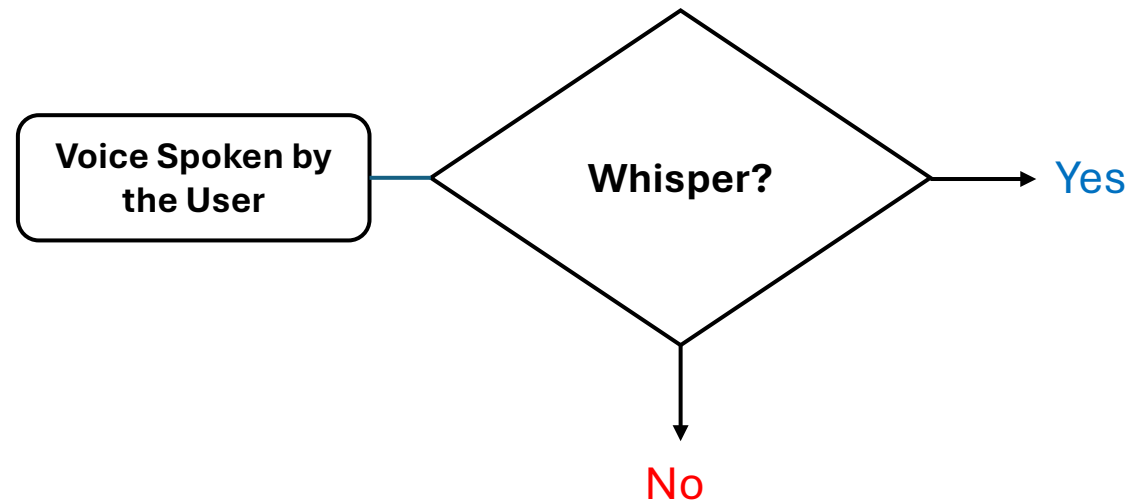
19 Q Yeah. So in the context of determining
20 whether the user is whispering or not, if the user
21 is, for example, at distance X versus at distance
1 MR. REEVES: Objection. Form. Incomplete
2 hypothetical.
3 A Yeah, I don't know. The patent does not
4 address that hypothetical situation directly. It
5 only considers independently and in isolation
6 classification based on distance and detection of
7 whispering.

Ex. 2020, 48:19-49:7; POR, 6

“Determine ... When the Voice Spoken by the User is a Whisper”

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.

Ex. 1001, 19 (cl.5)



POR, 7-8; POSR, 4-5

“Determine ... When the Voice Spoken by the User is a Whisper”

The Specification describes determining that the voice is a whisper based on a peak frequency threshold

- “Whisper” is defined as “an unvoiced voice spoken without vibrating the vocal chords”

be a whisper. A whisper is an unvoiced voice spoken without vibrating the vocal chords, but it is not limited to the unvoiced voice spoken without vibrating the vocal chords as long as being classified to the first voice V1A by, for example, a method described later.

For example, the voice classifying unit 42 determines the voice as a whisper to classify to the first voice V1A when the peak frequency is equal to or lower than the threshold, and determines the voice as not a whisper to classify to the second voice V1B when the peak frequency is larger than the threshold. Note that the voice classifying unit 42 can

Ex. 1001, 10:4-10:19; POR, 7-8; POSR, 4-5

Lack of a Factual Record

The Petition relies solely on *ipse dixit* expert statements to allege “voice classification”

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

Pet., 12

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

Ex. 1002, ¶60

No Factual Support

The Petition relies solely on *ipse dixit* expert statements to allege “voice classification”

Shimomura discloses that its voice classifying unit (processor) analyzes the voice to classify it as either a first or second voice based on the user-to-device distance. (Ex. 1002 ¶¶169-71.) Shimomura’s control unit includes a “speech synthesis unit” and a “speech character string transformation unit.” (Ex. 1004 ¶¶[0036]-[0037], [0046], [0058], [0067], Fig. 4.) These units may “transform the character string” of the device’s spoken content “according to the distance to the conversation partner.” (*Id.* ¶¶[0058], [0067]-[0070]; *id.* ¶¶[0060]-[0063], [0073]-[0084], Abstract, claims 1-24.) To do so, Shimomura’s control unit first classifies the voice.

(Ex. 1002 ¶169.)

Pet., 49-50

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

Ex. 1002, ¶169

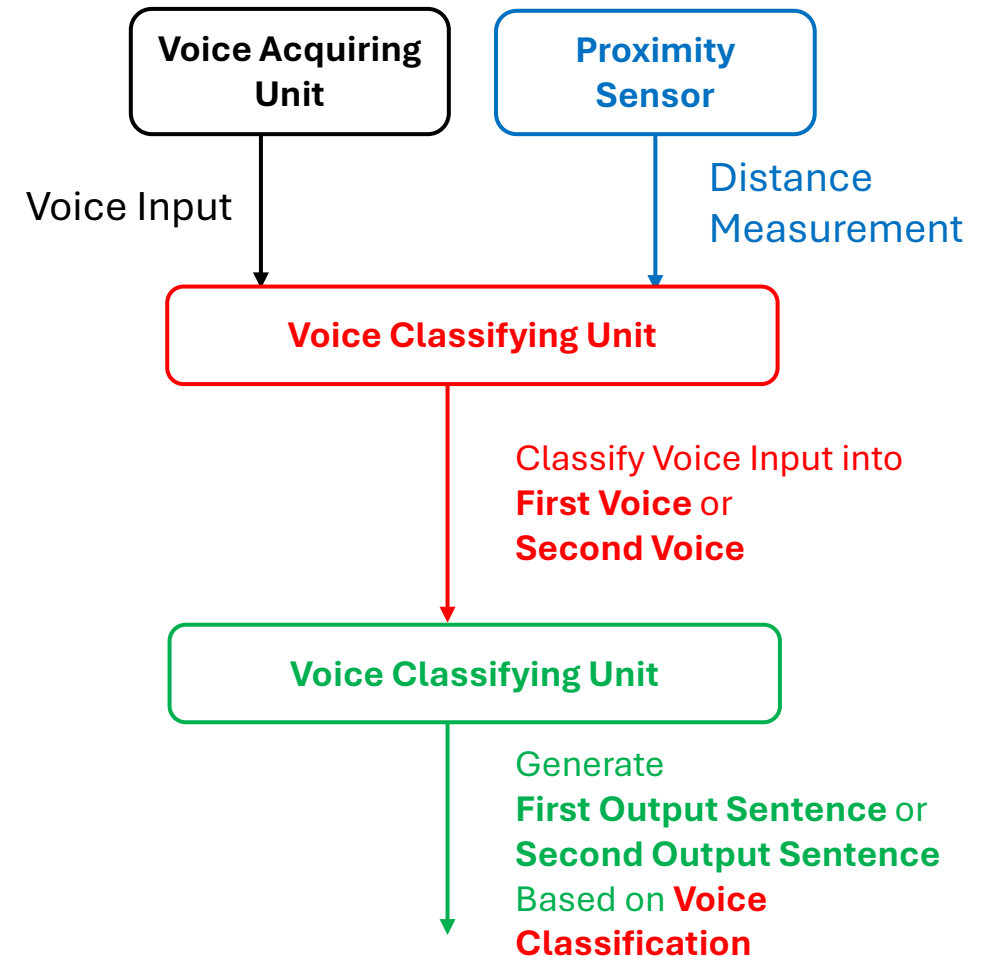
Petitioners Conflate Distance Measurements with Voice Classification

Claim Language:

1. An output-content control device comprising:

- [a] a voice acquiring unit ...;
- [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 ...;
- [d] a notification-information acquiring unit ...; and
- [e] an **output-content generating unit** configured to generate an output sentence

POR, 7-8; POSR, 5-8



Prior Art:
Shin

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

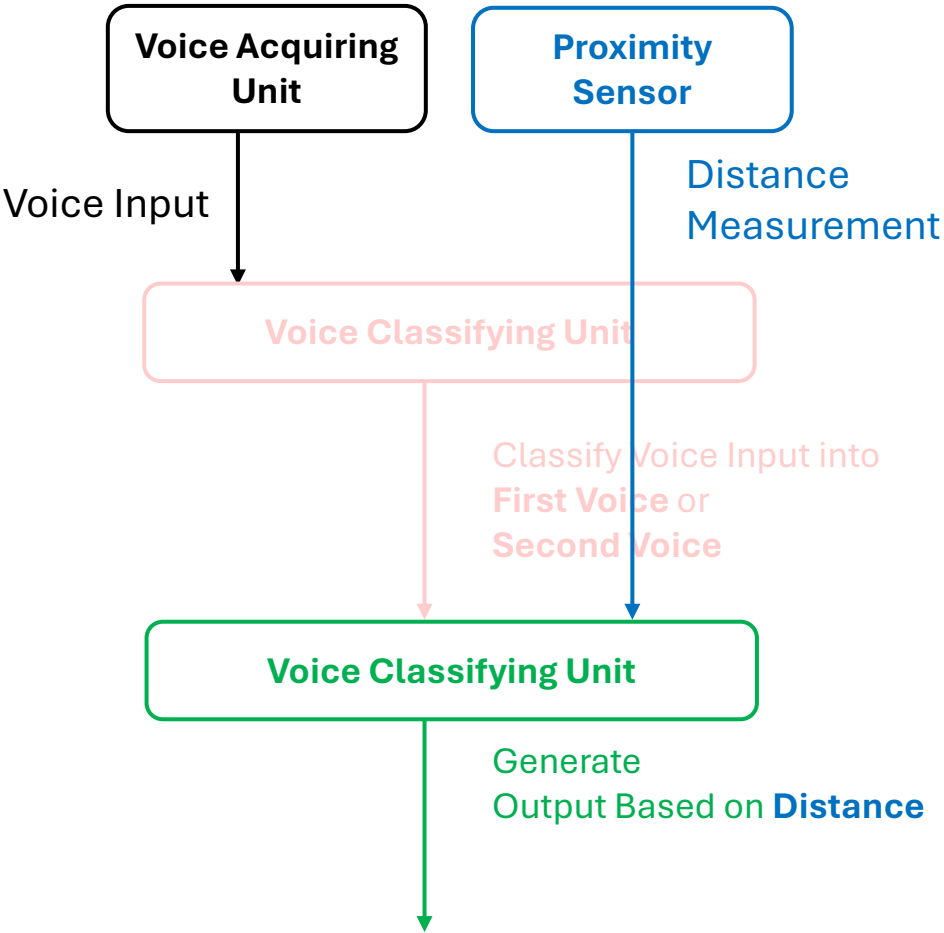
Shimomura
[FIG. 12]

Less than 200 cm	Less than 200 cm
None	Raise

55

FIG. 12 Phrase-end emphasis determination table

Ex. 1004, Fig. 12

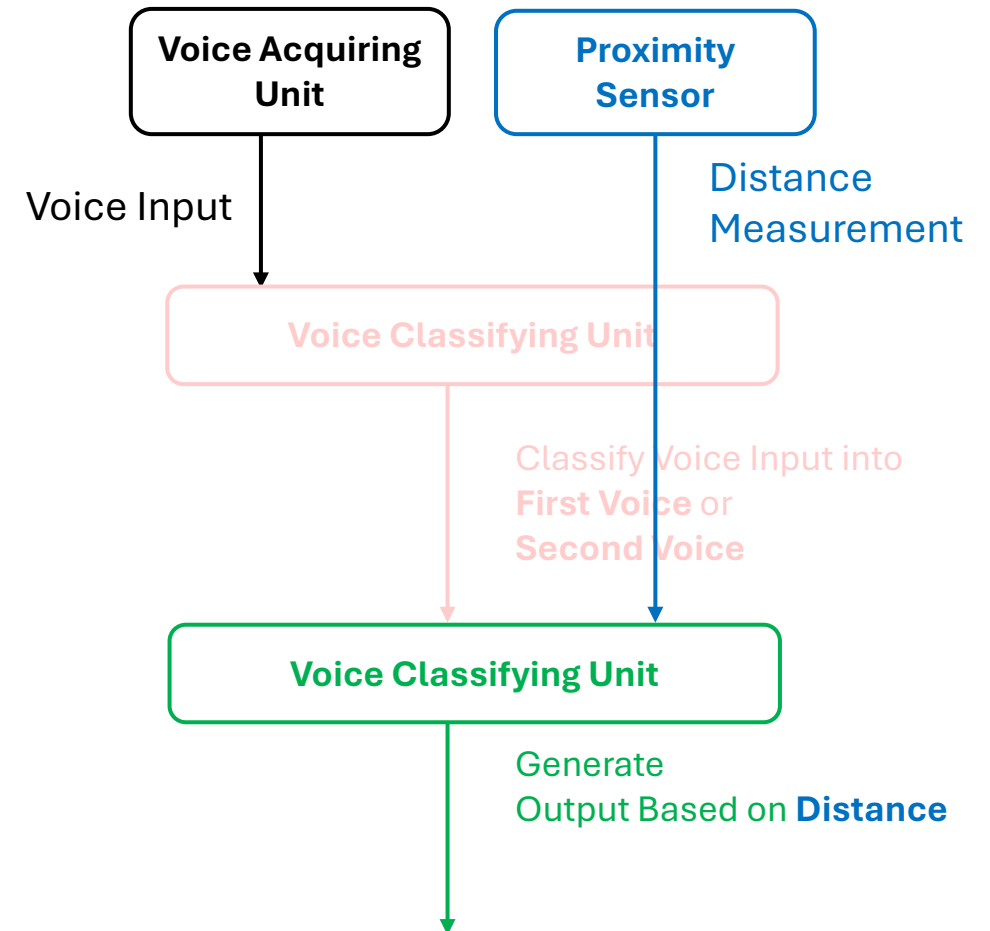


POR, 7-8; POSR, 5-8

Petitioners Conflate Distance Measurements with Voice Classification

Hypothetical Claim:

1. An output-content control device comprising:
 - [a] a voice acquiring unit ...; and
 - [b] **an output-content generating unit** configured to generate one of two output sentences based upon a distance between the user and the output-content control device by a **proximity sensor**.



POR, 7-8; POSR, 5-8

Prior Art Discloses Distance Classification, Not Voice Classification

Shin

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

Shimomura

[FIG. 12]

Less than 200 cm	Less than 200 cm
None	Raise

Input Distance

Output Content

55

FIG. 12 Phrase-end emphasis determination table

Ex. 1004, Fig. 12

POSR, 12-15

- Not even attempt at inherency
- *Ipsissimis verbis* strawman for failure of inherency. *Akzo N.V. v. United States Int'l Trade Comm'n*, 808 F.2d 1471, 1479 (Fed. Cir. 1986) (a reference disclosing “use of sulfuric acid” did not disclose “use of at least 98% concentrated sulfuric acid” because the challenger failed to establish inherency; not because the ITC applied an “*ipsissimis verbis* test.”)
- *Ipsissimis verbis* is a word for word concern.
- Does not apply where entire limitations are missing

Reply, 11, 21

Prior Art Does Not Disclose Word Replacement

Shin Discloses Abbreviation by Excluding (Not Replacing) Words

Original Content:

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

Abbreviated Content:

First Level – (1) + (2) + (3) + (4)

Second Level – (1) + (2) + (4)

Third Level – (1) + (2)

Fourth Level – (2)

(④). Prepare your umbrella when you go out. (⑤)” In this case, the content of a first abbreviation level may include information of ①, ②, ③, and ④ of the original content while excluding ⑤, and the content of a second abbreviation level may include information of ①, ② and ④ while excluding ③ and ⑤. Furthermore, the content of a third abbreviation level may only include information of ① and ②, and the content of a fourth abbreviation level (abstract content) may only include information of ②. As such, an output amount of information of content may vary or be adjusted by reconfiguring a portion of the original content that describes the content most precisely.

Ex. 1003, [0088]; POR, 28-30

Uncontested that Shimomura Does Not Disclose Word Replacement

Thus, Shimomura discloses adjusting the output content based on the user-to-device distance, for example to increase user comprehension. (Ex. 1002 ¶¶183-85.)

As discussed above, Aoyama discloses replacing words in output content, for example to adjust how names are presented. (*Supra* §§VII.A.6-VII.A.7.)

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

Pet., 55; Reply, 21-23; POR, 46; POSR, 19

Aoyama Does Not Disclose Replacing Words Based on Voice Classification

Aoyama discloses generating **responses** based on user's **speech style**

Original Content Taro Yamada

User's Speech Style Mr. Yamada

Generated Response Mr. Yamada

Ex. 1010, [0086]-[0089], [0121]; POR, 30

Aoyama discloses changing the **amount of information** output based on user's degree of **composure**

User's Degree Of Composure	Output
Calm	Response Only
Hurried	Response + Additional information

Ex. 1010, [0086]-[0089], [0121]-[0124]; POR, 28-30



Iwase Does Not Disclose Replacing Words Based on Voice Classification

Iwase discloses generating responses based on user attributes

- **User Attributes:** age, distinction of sex, language, and hometown

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; POR, 54

That attribute information includes “age, distinction of sex, language, and hometown.” *Id.*, 12:27-34. Iwase does not disclose any operation based on voice classification. It discloses adjusting outputs based on an “attention acquisition difficulty level,” which is calculated based on “noise level and a situation of the user,” as well as “attribute information of a user.” *Id.*, 10:35-67, 12:27-42, 15:41-45, Fig. 7; EX-2016, ¶147. “Attributed information of a user” in Iwase includes, for example, age, distinction of sex, language, and hometown.” EX-1017, 12:26-31. None of that involves any voice analysis, let alone voice classification. EX-2016, ¶147.

POR, 54

Prior Art Does Not Disclose Determining a Whisper

Grounds 1-3 are based on erroneous interpretation that claim 5 does not require determining a whisper:

PO argues that claim 5 requires determining that the voice is a whisper. (POR, 36.) It does not. Instead, claim 5 requires that the device be configured to determine that “the voice” is “the predetermined voice” in circumstances where the user whispers to the device. (EX-1001, claim 5; EX-1202 ¶19.) Nothing in the claim requires the device to *determine* whether the received voice is a whisper.

Reply, 16-17

Schuster discloses determining whisper for purpose of switching modes

Once the device has detected an intended private interaction, this mode change alters the way in which the device interacts with the user. For example, the device may typically play audible speech at a normal spoken volume back to the user after the user makes a query. However, if private interaction intention is detected, the device may significantly lower the volume of the speech feedback so that only the user or someone very close to the device (e.g., within one or two feet) will be able to hear the response.

EX. 1016, 3:31-39

From the normal mode state **201**, the device will move to the private mode state **203** in an embodiment when the device detects an intended private interaction from the user.

EX. 1016, 5:37-39

POR, 57-59

No Motivation to Combine

Schuster

- *Detects whisper based on microphone signal analysis*
- No guidance on how to detect whisper based on distance

Shin/Shimomura

- Allegedly *classify voices based on proximity sensor measurements*

However, as the Petition acknowledges, “Schuster’s device detects whispered input through *microphone signal analysis*.” *Id.*, 65-66 (citing EX-1016, 3:2-4). Schuster offers no guidance whatsoever on how a whisper could be detected based on distance at all—or why a POSITA would have been motivated to do so. Quite the opposite. Introducing distance would risk confounding the determination. EX-2017, 56:15-57:5 A POSITA would have recognized that Schuster’s microphone signal analysis involves entirely different signals, metrics, and calculations than those in Shin and Shimomura’s proximity sensor calculation. *See, e.g.*, EX-1016, 3:1-11, 6:30-41.

Ex. 2016, ¶158; POR, 58-59

Obviousness

Obviousness

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

Sixth, such a configuration applies a known technique (Aoyama's word replacement) to a known device and method (Shin's content tailoring system) that is ready for improvement and yields predictable results (providing content to distant users that is tailored using word replacement to address privacy concerns). (*Id.*)

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

Pet. 27; POR 61-62; POPR 5-8

Third, such a configuration reflects applying a known technique (Iwase's word replacement) to improve a similar device and method (Shin's content output system) in the same way (providing easier-to-grasp output). (*Id.*)

Fourth, such a configuration also applies a known technique (Iwase's word replacement) to a known device and method (Shin's content abbreviation system) that is ready for improvement and yields predictable results (delivering content to users at greater distances that is abbreviated using word replacement). (*Id.*)

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

Pet. 43; POR 61-62; POPR 5-8

Fifth, such a configuration applies a known technique ([reference A disclosure]) to improve a similar device and method ([reference B disclosure]) in the same way.

Pet. 27

In car engines without computer-controlled throttles, the accelerator pedal interacts with the throttle via cable or other mechanical link. The pedal arm acts as a lever rotating around a pivot point. In a cable-actuated throttle control the rotation caused by pushing down the pedal pulls a cable, which in turn pulls open valves in the carburetor or fuel injection unit. The wider the valves open, the more fuel and air are released, causing combustion to increase and the car to accelerate. When the driver takes his foot off the pedal, the opposite occurs as the cable is released and the valves slide closed.

In the 1990's it became more common to install computers in cars to control engine operation. Computer-controlled [*408] throttles open and close valves in response to electronic signals, not through force transferred from the pedal by a mechanical link. Constant, delicate adjustments of air and fuel mixture are possible. The computer's rapid processing of factors beyond the pedal's position improves fuel efficiency and engine performance.

For a computer-controlled throttle to respond to a driver's operation of the car, the computer must know what is happening with the pedal. A cable or mechanical link does not suffice for this purpose; at some point, an electronic sensor is necessary to translate the mechanical operation into digital data the computer can understand.

Before discussing sensors further we turn to the mechanical design of the pedal itself. In the traditional design a pedal can be pushed down or released but cannot have its position in the footwell adjusted by sliding the pedal forward or back. As a result, a driver who wishes to be closer or farther from the pedal must either reposition himself in the driver's seat or move the seat in some way. In cars with deep footwells these are imperfect solutions for drivers of smaller stature. To solve the problem, inventors, beginning in the 1970's, designed pedals that could be adjusted to change their location in the footwell. Important for this case are two adjustable pedals disclosed in U.S. Patent Nos. 5,010,782 (filed July 28, 1989) (Asano) and 5,460,061 (filed Sept. 17, 1993) (Redding). The Asano patent reveals a support structure that houses the pedal so that even when the pedal location is adjusted relative to the driver, one of the pedal's pivot points stays fixed. The pedal is also designed so that the force necessary to push the pedal down is the same regardless of adjustments to its location. The Redding patent reveals a different, sliding mechanism where both the pedal and the pivot point are adjusted.

We return to sensors. Well before Engelgau applied for his challenged patent, some inventors had obtained patents involving electronic pedal sensors for computer-controlled [*409] throttles. These inventions, such as the device disclosed in U.S. Patent No. 5,241,936 (filed Sept. 9, 1991) ('936), taught that it was preferable to detect the pedal's position in the pedal assembly, not in the engine. The '936 patent disclosed a pedal with an electronic sensor on a pivot point in the pedal assembly. U.S. Patent No. 5,063,811 (filed July 9, 1990) (Smith) taught that to prevent the wires connecting the sensor to the computer from chafing and wearing out, and to avoid grime and damage from the driver's foot, the sensor should be put on a fixed part of the pedal assembly rather than in or on the pedal's footpad.

In addition to patents for pedals with integrated sensors inventors obtained patents for self-contained modular sensors. A modular sensor is designed independently of a given pedal so that it can be taken off the shelf and attached to mechanical pedals of various sorts, enabling the pedals to be used in automobiles with computer-controlled throttles. One such sensor was disclosed in U.S. Patent No. 5,385,068 (filed Dec. 18, 1992) ('068). In 1994, Chevrolet manufactured a line of trucks using modular sensors "attached to the pedal assembly support bracket, adjacent to the pedal and engaged with the pivot shaft about which the pedal rotates in operation." 298 F. Supp. 2d 581, 589 (ED Mich. 2003).

The prior art contained patents involving the placement of sensors on adjustable pedals as well. For example, U.S. Patent No. 5,819,593 (filed Aug. 17, 1995) (Rixon) discloses an adjustable pedal assembly with an electronic sensor for detecting the pedal's position. In the Rixon pedal the sensor is located in the pedal footpad. The Rixon pedal was known to suffer from wire chafing when the pedal was depressed and released.

This short account of pedal and sensor technology leads to the instant case.

There then existed a marketplace that created a strong incentive to convert mechanical pedals to electronic pedals, and the prior art taught a number of methods for achieving this advance. The Court of Appeals considered the issue too narrowly by, in effect, asking whether a pedal designer writing on a blank slate would have chosen both Asano and a modular sensor similar to the ones used in the Chevrolet truckline and disclosed in the '068 patent. The District Court employed this narrow inquiry as well, though it reached the correct result nevertheless. The proper question to have asked was whether a pedal designer of ordinary skill, facing the wide range of needs created by developments in the field of endeavor, would have seen a benefit to upgrading Asano with a sensor.

In automotive design, as in many other fields, the interaction of multiple components means that changing one component often requires the others to be modified as well. Technological developments made it clear that engines using computer-controlled throttles would become standard. As a result, designers might have decided to design new pedals from scratch; but they also would have had reason to make pre-existing pedals work with the new engines. Indeed, upgrading its own pre-existing model led KSR to design the pedal now accused of infringing the Engelgau patent.

For a designer starting with Asano, the question was where to attach the sensor. The consequent legal question, then, is whether a pedal designer of ordinary skill starting with Asano would have found it obvious to put the sensor on [*425] a fixed pivot point. The prior art discussed above leads us to the conclusion that attaching the sensor where both KSR and Engelgau put it would have been obvious to a person of ordinary skill.

The '936 patent taught the utility of putting the sensor on the pedal device, not in the engine. Smith, in turn, explained to put the sensor not on the pedal's footpad but instead on its support structure. And from the known wire-chafing problems of Rixon, and Smith's teaching that "the pedal assemblies must not precipitate any motion in the connecting wires," Smith, col. 1, ll. 35-37, Supp. App. 274, the designer would know to place the sensor on a nonmoving part of the pedal structure. The most obvious nonmoving point on the structure from which a sensor can easily detect the pedal's position is a pivot point. The designer, accordingly, would follow Smith in mounting the sensor on a pivot, thereby designing an adjustable electronic pedal covered by claim 4.

Just as it was possible to begin with the objective to upgrade Asano to work with a computer-controlled throttle, so too was it possible to take an adjustable electronic pedal like Rixon and seek an improvement that would avoid the wire-chafing problem. Following similar steps to those just explained, a designer would learn from Smith to avoid sensor movement and would come, thereby, to Asano because Asano disclosed an adjustable pedal with a fixed pivot.

Teleflex indirectly argues that the prior art taught away from attaching a sensor to Asano because Asano in its view is bulky, complex, and expensive. The only evidence Teleflex marshals in support of this argument, however, is the Radcliffe declaration, which merely indicates that Asano would not have solved Engelgau's goal of making a small, simple, and inexpensive pedal. What the declaration does not indicate is that Asano was somehow so flawed that there was no reason to upgrade it, or pedals like it, to be compatible with modern engines. Indeed, Teleflex's own declarations [*426] refute this conclusion. Dr. Radcliffe states that Rixon suffered from the same bulk and complexity as did Asano. See id., at 206. Teleflex's other expert, however, explained that Rixon was itself designed by adding a sensor to a pre-existing mechanical pedal. See id., at 209. If Rixon's base pedal was not too flawed to upgrade, then Dr. Radcliffe's declaration does not show Asano was either. Teleflex may have made a plausible argument that Asano is inefficient as compared to Engelgau's preferred embodiment, but to judge Asano against Engelgau would be to engage in the very hindsight bias Teleflex rightly urges must be avoided. Accordingly, Teleflex has not shown anything in the prior art that taught away from the use of Asano.

KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 407-09, 424-26 (2007)

Illogical Combination

- Tables in Shin and Shimomura relates only to making a discrete distance measurement
- Shin and Shimomura each operates the same regardless of voice change
- Modifying Shin and Shimomura to conduct voice analyses would have made the devices more complex, not simpler

POR, 60-61

Failure to Apply Second *Graham* Factor



A determination of whether a patent claim is invalid as obvious under § 103 requires consideration of **all four** Graham factors, and it is error to reach a conclusion of obviousness until all of those factors are considered.

Artic Cat Inv. v. Bombardier Rec. Prods., 876 F.3d 1350, 1358(Fed. Cir. 2017) ; Sur-Reply, 21-22

Improperly Articulated Combinations for Ground 5

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

As discussed above, the references in Grounds 1a-4 disclose each limitation of claims 1-7. (*Supra* §§VII-XII.) The claims also would have been obvious over those references and Kristjansson.

Pet., 67; POR, 62