

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.,
Petitioner,

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner

Case No. IPR2025-00565 (Patent No. 11,069,337)

Patent Owner's Demonstratives

Introduction and Background

Grounds

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

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

Pet., 5-6

Challenged Patent

(54) VOICE-CONTENT CONTROL DEVICE, VOICE-CONTENT CONTROL METHOD, AND NON-TRANSITORY STORAGE MEDIUM

US011069337B2

(12) United States Patent
Naganuma

(10) Patent No.: US 11,069,337
(45) Date of Patent: Jul. 1, 2020

(54) VOICE-CONTENT CONTROL DEVICE,
VOICE-CONTENT CONTROL METHOD,
AND NON-TRANSITORY STORAGE
MEDIUM

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

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

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

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 108 days.

(21) Appl. No.: 16/290,983

(22) Filed: Mar. 4, 2019

(65) Prior Publication Data
US 2019/0279611 A1 Sep. 12, 2019

(30) Foreign Application Priority Data
Mar. 6, 2018 (JP) JP2018-039754

(51) Int. Cl.
G10L 13/08 (2013.01); G06F 3/167
G10L 17/26 (2013.01)
(Continued)

(52) U.S. Cl.
CPC G10L 13/08 (2013.01); G06F 3/167
(2013.01); G10L 15/08 (2013.01); G10L 15/16
(2013.01); G10L 15/18 (2013.01); G10L
15/1807 (2013.01); G10L 15/22 (2013.01);
G10L 17/26 (2013.01); G10L 15/783
(2013.01);
(Continued)

(58) Field of Classification Search
CPC G06N 3/08; G06N 20/00; G06N 3/02; G10L

(56) References Cited
U.S. PATENT DOCUMENTS
5,852,804 A 12/1998 Sato
2013/0325471 A1* 12/2013 Rachevsky G06N 20/00
704,244
(Continued)

FOREIGN PATENT DOCUMENTS
JP 04-204700 7/1992
Primary Examiner — Linda Wong
(74) Attorney, Agent, or Firm — Amin, Turocy & Watson,
LLP

(57) ABSTRACT
A voice-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 classify the voice as
either one of a first voice or a second voice, a process
executing unit configured to analyze the acquired voice to
execute processing required by the user, and a voice-content

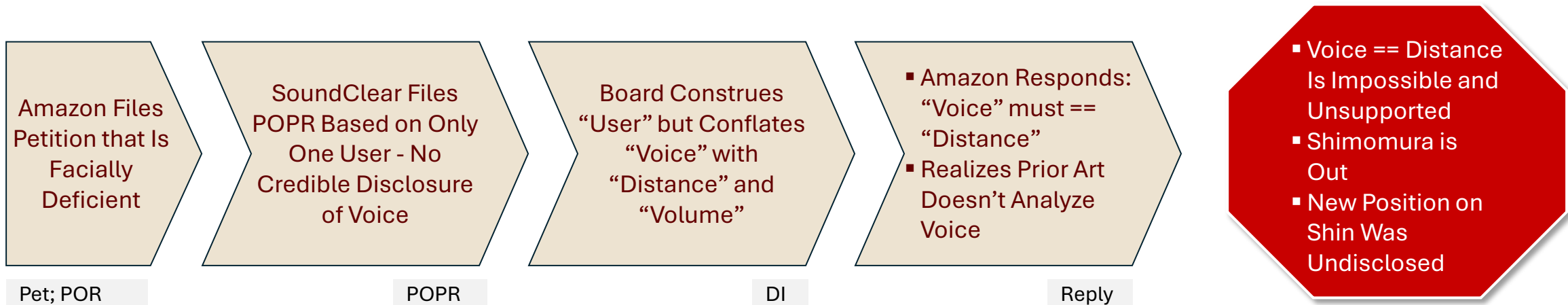
5 Claims, 6 Drawing Sheets

```
graph TD
    subgraph Device
        direction TB
        VCU[VOICE-CONTENT CONTROL UNIT]
        VACU[VOICE ACQUIRING UNIT]
        VCCU[VOICE CLASSIFYING UNIT]
        PEU[PROCESS EXECUTING UNIT]
        VCCU2[VOICE-CONTENT CONTROL UNIT]
        VCCU3[VOICE-CONTENT CONTROL UNIT]
        VCCU4[VOICE-CONTENT CONTROL UNIT]
        VCCU5[VOICE-CONTENT CONTROL UNIT]
        VCCU6[VOICE-CONTENT CONTROL UNIT]
        VCCU7[VOICE-CONTENT CONTROL UNIT]
        VCCU8[VOICE-CONTENT CONTROL UNIT]
        VCCU9[VOICE-CONTENT CONTROL UNIT]
        VCCU10[VOICE-CONTENT CONTROL UNIT]
        VCCU11[VOICE-CONTENT CONTROL UNIT]
        VCCU12[VOICE-CONTENT CONTROL UNIT]
        VCCU13[VOICE-CONTENT CONTROL UNIT]
        VCCU14[VOICE-CONTENT CONTROL UNIT]
        VCCU15[VOICE-CONTENT CONTROL UNIT]
        VCCU16[VOICE-CONTENT CONTROL UNIT]
        VCCU17[VOICE-CONTENT CONTROL UNIT]
        VCCU18[VOICE-CONTENT CONTROL UNIT]
        VCCU19[VOICE-CONTENT CONTROL UNIT]
        VCCU20[VOICE-CONTENT CONTROL UNIT]
        VCCU21[VOICE-CONTENT CONTROL UNIT]
        VCCU22[VOICE-CONTENT CONTROL UNIT]
        VCCU23[VOICE-CONTENT CONTROL UNIT]
        VCCU24[VOICE-CONTENT CONTROL UNIT]
        VCCU25[VOICE-CONTENT CONTROL UNIT]
        VCCU26[VOICE-CONTENT CONTROL UNIT]
        VCCU27[VOICE-CONTENT CONTROL UNIT]
        VCCU28[VOICE-CONTENT CONTROL UNIT]
        VCCU29[VOICE-CONTENT CONTROL UNIT]
        VCCU30[VOICE-CONTENT CONTROL UNIT]
        VCCU31[VOICE-CONTENT CONTROL UNIT]
        VCCU32[VOICE-CONTENT CONTROL UNIT]
        VCCU33[VOICE-CONTENT CONTROL UNIT]
        VCCU34[VOICE-CONTENT CONTROL UNIT]
        VCCU35[VOICE-CONTENT CONTROL UNIT]
        VCCU36[VOICE-CONTENT CONTROL UNIT]
        VCCU37[VOICE-CONTENT CONTROL UNIT]
        VCCU38[VOICE-CONTENT CONTROL UNIT]
        VCCU39[VOICE-CONTENT CONTROL UNIT]
        VCCU40[VOICE-CONTENT CONTROL UNIT]
        VCCU41[VOICE-CONTENT CONTROL UNIT]
        VCCU42[VOICE-CONTENT CONTROL UNIT]
        VCCU43[VOICE-CONTENT CONTROL UNIT]
        VCCU44[VOICE-CONTENT CONTROL UNIT]
        VCCU45[VOICE-CONTENT CONTROL UNIT]
        VCCU46[VOICE-CONTENT CONTROL UNIT]
        VCCU47[VOICE-CONTENT CONTROL UNIT]
        VCCU48[VOICE-CONTENT CONTROL UNIT]
        VCCU49[VOICE-CONTENT CONTROL UNIT]
        VCCU50[VOICE-CONTENT CONTROL UNIT]
        VCCU51[VOICE-CONTENT CONTROL UNIT]
        VCCU52[VOICE-CONTENT CONTROL UNIT]
        VCCU53[VOICE-CONTENT CONTROL UNIT]
        VCCU54[VOICE-CONTENT CONTROL UNIT]
        VCCU55[VOICE-CONTENT CONTROL UNIT]
        VCCU56[VOICE-CONTENT CONTROL UNIT]
        VCCU57[VOICE-CONTENT CONTROL UNIT]
        VCCU58[VOICE-CONTENT CONTROL UNIT]
        VCCU59[VOICE-CONTENT CONTROL UNIT]
        VCCU60[VOICE-CONTENT CONTROL UNIT]
        VCCU61[VOICE-CONTENT CONTROL UNIT]
        VCCU62[VOICE-CONTENT CONTROL UNIT]
        VCCU63[VOICE-CONTENT CONTROL UNIT]
        VCCU64[VOICE-CONTENT CONTROL UNIT]
        VCCU65[VOICE-CONTENT CONTROL UNIT]
        VCCU66[VOICE-CONTENT CONTROL UNIT]
        VCCU67[VOICE-CONTENT CONTROL UNIT]
        VCCU68[VOICE-CONTENT CONTROL UNIT]
        VCCU69[VOICE-CONTENT CONTROL UNIT]
        VCCU70[VOICE-CONTENT CONTROL UNIT]
        VCCU71[VOICE-CONTENT CONTROL UNIT]
        VCCU72[VOICE-CONTENT CONTROL UNIT]
        VCCU73[VOICE-CONTENT CONTROL UNIT]
        VCCU74[VOICE-CONTENT CONTROL UNIT]
        VCCU75[VOICE-CONTENT CONTROL UNIT]
        VCCU76[VOICE-CONTENT CONTROL UNIT]
        VCCU77[VOICE-CONTENT CONTROL UNIT]
        VCCU78[VOICE-CONTENT CONTROL UNIT]
        VCCU79[VOICE-CONTENT CONTROL UNIT]
        VCCU80[VOICE-CONTENT CONTROL UNIT]
        VCCU81[VOICE-CONTENT CONTROL UNIT]
        VCCU82[VOICE-CONTENT CONTROL UNIT]
        VCCU83[VOICE-CONTENT CONTROL UNIT]
        VCCU84[VOICE-CONTENT CONTROL UNIT]
        VCCU85[VOICE-CONTENT CONTROL UNIT]
        VCCU86[VOICE-CONTENT CONTROL UNIT]
        VCCU87[VOICE-CONTENT CONTROL UNIT]
        VCCU88[VOICE-CONTENT CONTROL UNIT]
        VCCU89[VOICE-CONTENT CONTROL UNIT]
        VCCU90[VOICE-CONTENT CONTROL UNIT]
        VCCU91[VOICE-CONTENT CONTROL UNIT]
        VCCU92[VOICE-CONTENT CONTROL UNIT]
        VCCU93[VOICE-CONTENT CONTROL UNIT]
        VCCU94[VOICE-CONTENT CONTROL UNIT]
        VCCU95[VOICE-CONTENT CONTROL UNIT]
        VCCU96[VOICE-CONTENT CONTROL UNIT]
        VCCU97[VOICE-CONTENT CONTROL UNIT]
        VCCU98[VOICE-CONTENT CONTROL UNIT]
        VCCU99[VOICE-CONTENT CONTROL UNIT]
        VCCU100[VOICE-CONTENT CONTROL UNIT]
    end
```

(57) ABSTRACT

A voice-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 classify the voice as either one of a first voice or a second voice, a process executing unit configured to analyze the acquired voice to execute processing required by the user, and a voice-content

How We Got Here



IPR Petition – Patent 11,069,337
Amazon.com, Inc., et al. v. SoundClear Technologies LLC

VI. CLAIM CONSTRUCTION

No claim terms require construction to resolve the invalidity challenges here.

Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd., 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999). For purposes of this proceeding only, Petitioner assumes the claims are not invalid under §112.

Pet., 8

Expert Record

Transcript of Richard Martin Stern, Jr., Ph.D. Conducted on October 30, 2025

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.

IN RE: Petitioners : IPR2025-00565
v. : U.S. Patent
SOUNDCLEAR TECHNOLOGIES LLC, : No. 11,069,337
Patent Owner. :
----- x

Videotaped Deposition of
RICHARD MARTIN STERN, JR., Ph.D.
Conducted Remotely
Thursday, October 30, 2025
12:15 p.m. Eastern Standard Time

Job No. 604599
Pages: 1 - 119
Reported by: Janet A. Hamilton, RDR

EX-2015, 54:1-6; POR, 7

12 Q Can you looking at table 3, can you --
13 does it provide any information regarding the
14 user's voice in table 3?
15 MR. REEVES: Objection. Form. Asked and
16 answered.
17 A It has been asked and answered. I'm not
18 sure what more I can tell you that I haven't
19 already.
20 Q Well, you keep -- you keep answering in
21 the context of classification. That's not my
22 question, sir. My question is simply, does table
1 3 provide any information regarding the user's
2 voice, not whether you can use table 3 to classify
3 the voice?
4 MR. REEVES: Same objections.
5 A No. Just so we don't continue iterating
6 this for the rest of the day, it tells you about
7 the distance between the user and the electronic
8 device and nothing else.

EX-2015, 72:12-73:8; POR, 27-28

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

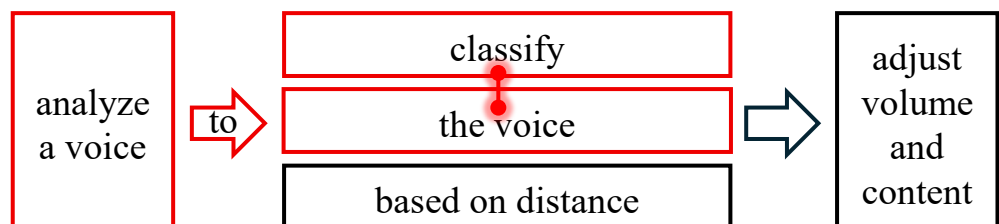
22 Would a POSITA have been able to determine
1 how a device classifies inputs simply by looking at
2 how the device adjusts outputs?
3 MR. REEVES: Objection. Form. Incomplete
4 hypothetical. Outside the scope.
5 THE WITNESS: I think, in general, the
6 answer would be no. I don't believe that that's
7 relevant to the claims of the present patent or the
8 '337 patent.

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

EX-2016, 62:22-63:8

Classify Voice

- A device that classifies “voices” based on specific criteria



- *Hypothetical Claim:* A device that measures distance and adjusts output volume

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, 6-21; POSR, 3-5

1[pre]. 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
- [f] the **voice-content** generating unit is further configured to
generate a first output sentence as the output sentence when the acquired **voice** has been classified as the first **voice**, and
- [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.

EX-1001; POR, 12-13

1[pre]. 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

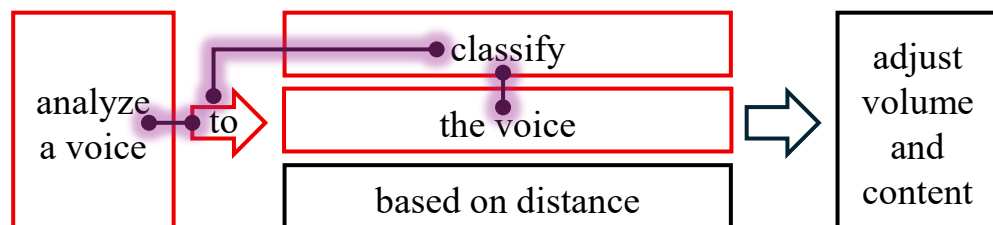
- [e] an output controller configured to adjust a **sound volume** of voice data obtained by converting the output **sentence** therein;
- [f] the voice-content generating unit is further configured to generate a first output **sentence** as the output **sentence** when the acquired voice has been classified as the first voice, and
- [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** therein differs from the **sound volume** of the voice **data** obtained by converting the second output **sentence** therein.

EX-1001; POR, 12-13

I. Preliminary Construction Went Astray

It is “unjust to the public, as well as an evasion of the law, to construe [an invention] in a manner different from the plain import of its terms.”

Phillips v. AWH Corp., 415 F.3d 1303, 1312 (Fed. Cir. 2005).



- [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;

EX-1001; POR, 12-13; POSR, 2, 5-19

Preliminary Construction

These controlling case authorities are applicable here. At this stage, we preliminarily construe the scope of the claim 1[b] term “a user” as covering one or more users. Likewise, we preliminarily construe the scope of the 1[b] phrase “a voice spoken by a user” as covering one or more voices spoken by a single user, or by plural users.

Accordingly, for the purpose of our decision on institution, our preliminary claim construction only requires Petitioner to show a single user who may perform both a first and second voice depending upon the single user’s changing distance (and associated changing voice volume) from the voice-content control device.

COURTENAY, *Administrative Patent Judge*.

DECISION
Granting Institution of *Inter Partes* Review
35 U.S.C. § 314

DI, 16; see also p. 24 (discussing construction in terms of argument regarding input/output distinction)

The Parties Agree that the Preliminary Construction Went Astray

“a user”

- parties and board agree that it covers one or more user

Remaining Portions of Construction

- parties agree that it is incorrect

Affect on “Classify”

- A discrete concept
- Dr. Stern agrees

POR, 6; POSR, 3-5

C. “Voice”

In the Institution Decision, the Board stated that a single user’s voice may be classified “depending upon the single user’s changing distance (*and associated changing voice volume*) from the voice-content control device.” (DI, 16 (emphasis added).) To the extent that the Board found the claims to *require* a change in distance *and volume*, Petitioner agrees with Patent Owner that this would be incorrect. The claims *require* only classification based on distance. If the distance threshold is 1m, a user’s voice may be classified as a “first voice” when the user-to-device distance is 0.5m and a “second voice” when the user-to-device distance is 1.5m, even if the volume of the received voice is the same (e.g., because the user spoke louder at 1.5m).

EX-2015, 28

Mr. Joseph McAlexander, P.E. (for Patent Owner)

Affect on “Classify”

- A discrete concept
- Dr. Stern agrees

POR, 6; POSR, 3-5

UNITED STATES PATENT AND TRADEMARK OFFICE

BE

43. A POSITA would have understood “classify” to require something discrete, not a mere measurement. A POSITA would not have understood “classify” to be used in the context of measuring a single, continuously changing variable (like volume or distance). A POSITA would have understood the claims of the '337 patent to require a classification, or distinction between, at least two separately identifiable voices.

EX-2013, ¶43; POR, 6



Joseph C. McAlexander III

PROFESSIONS, PUBLICATIONS, EDUCATION

PROFESSIONS AND AWARDS

Electronics Engineers, Inc. (IEEE), Senior Member. Societies: Solid State Circuits (ESS) Engineers Engineers, Registered License #79454 Who's Who in Executives and Businesses Who's Who in the Registry of Business Leaders

PUBLICATIONS

1- IEEE Transactions on Computers, Aug '83, pgs. 139-142, The Memory

2- Archives of Biochemistry and Biophysics, Dec'81, Vol. 212, No. 2, Equilibrium Constants under Physiological Conditions for the Reactions of D-3-Phosphoglycerate Dehydrogenase and L-Phosphoserine Aminotransferase.

3- International Electron Devices Meeting, Dec '79, pgs. 355-357, Sub 100ns 16K x 1 MOS Dynamic RAM Using a Grounded Substrate.

EDUCATION PROFILE

1980 - 1985 Taught Solid State Device Physics, Semiconductor Processing, Circuit Design Techniques, and Statistical Quality Control Methods

Effectiveness Training and Japanese Manufacturing Techniques, Participative Problem Solving courses

1975 - 1976 1.5 years Graduate Study in Neural Science, the University of Texas Graduate School of Biomedical Science

1965 - 1969 BSEE, North Carolina State University

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

DECLARATION OF JOSEPH MCALEXANDER

7

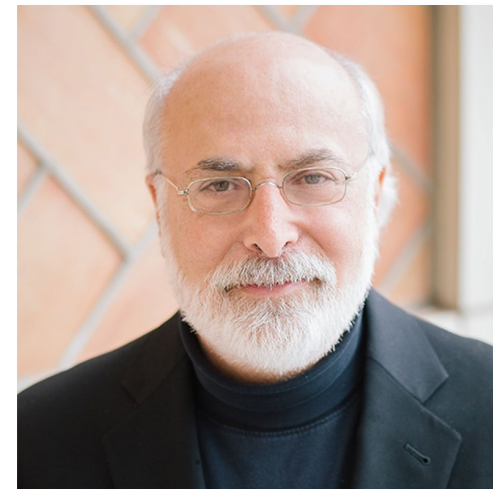
Dr. Richard Stern (for Petitioners)

Affect on “Classify”

- it is a discrete concept
- Dr. Stern agrees

EX-2016, 64:5-66:14; POSR, 6-7

5 Q I'd like to discuss a hypothetical
6 output-content control device that I came up with
7 this morning.
8 A We're always inventing.
9 Q I'm going to -- I'll call it a new
10 motivational output-content control device, and it's
11 pretty simple. When a user asks a question, it
12 measures the volume detected, and then it uses, you
13 know, a language model to develop an answer, and it
14 responds at twice the volume at which the user was
15 speaking.
16 So my logic would say output volume of the
17 device, call it OV, equals two times the input
18 volume of the device, IV.
19 In that hypothetical, is my device
20 **classifying** input voices?
21 MR. REEVES: Objection. Form. Calls for
22 speculation. Incomplete hypothetical. Outside the
1 scope.
2 THE WITNESS: **No.**



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

Output Volume (OV) = 2 × Input Volume (IV)

“Classify”

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

4 Q Why not?

5 MR. REEVES: Same objections.

6 THE WITNESS: Classifying the device, as I

7 see it, or classifying anything as I see it,

8 describes taking an input measurement that, in

9 general, is continuous in nature and producing an

10 output that's one of a discrete number of categories

11 in the patents in question. Is the speaker closer

12 or is the speaker farther?

13 The system that you described really is --

14 takes the input, measures the volume and outputs at

15 twice the volume, but there are no discrete

16 categories that are the results of the

17 classification decision.

18 BY MR. REYNOLDS:

19 Q What if I told you that my device processes

20 everything digitally? Would that change your

21 opinion?

22 A No.

1 Q Because when you do digital signal

2 processing, everything has to be -- all variables,

3 inputs and outputs, have to be discrete, right?

4 MR. REEVES: Objection. Form. Calls for

5 speculation. Outside the scope.

6 THE WITNESS: Well, no. I teach the DSP

7 course at Carnegie Mellon, so this is my home turf.

8 But, no, we're still talking about

9 regardless of whether the process is continuous

10 using conventional components or digital or discrete

11 time using a computer, we're still talking about,

12 presumably, unlimited number of continuous or --

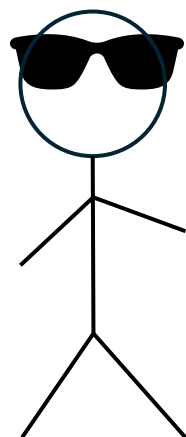
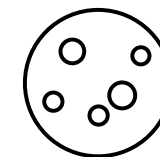
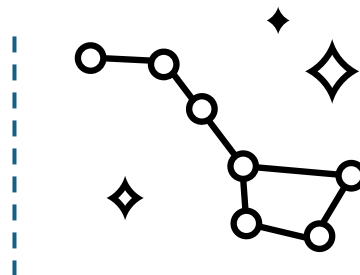
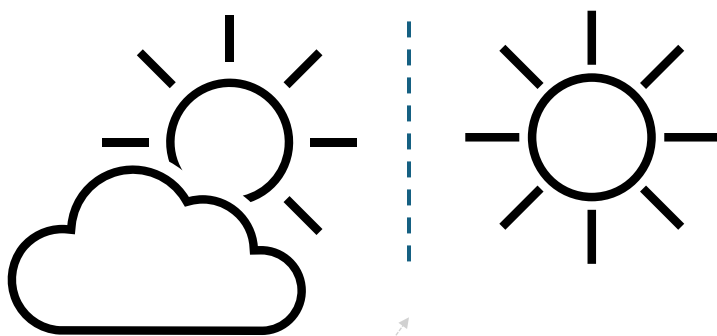
13 discrete, but unlimited number of possible

14 outputs/outcomes. That's not classification.

EX-2016, 64:5-66:14; POSR, 6-7

$$\text{Output Volume (OV)} = 2 \times \text{Input Volume (IV)}$$

Classification Versus Measuring



EX-2013, ¶53; POR, 16

53. The claimed classifying unit uses those characteristics to classify voices. Notably, each can be used alternatively or in different combinations. But no such changes alter the nature of what is being classified: voice. Certainly, a POSITA would not have understood use of a measurement to mean that any single characteristic defines the to-be-classified voice. That would be akin to stating that deciding how sunny it is outside is the same as classifying celestial objects—simply because measuring luminosity is one method that one might use to determine whether something in the sky is a planet or a star.

UNITED STATES PATENT AND TRADEMARK OFFICE

AND APPEAL BOARD

M, INC.,
VICES LLC, and
RVICES, INC.,
rs,

NOLOGIES LLC,
mer.

25-00565
11,069,337

PH MCALEXANDER

“Voice”

Affect on “Voice”

- heavily diminished in the claim
- redefined to be the same as “distance (and associated changing voice volume) from the content control device”

POR, 7-17; POSR, 3-5

These controlling case authorities are applicable here. At this stage, we preliminarily construe the scope of the claim 1[b] term “a user” as covering one or more users. Likewise, we preliminarily construe the scope of the 1[b] phrase “a voice spoken by a user” as covering one or more voices spoken by a single user, or by plural users.

Petitioner,

v.

Accordingly, for the purpose of our decision on institution, our preliminary claim construction only requires Petitioner to show a single user who may perform both a first and second voice depending upon the single user’s changing distance (and associated changing voice volume) from the voice-content control device.

COURTENAY, Administrative Patent Judge.

DECISION
Granting Institution of *Inter Partes* Review
35 U.S.C. § 314

DI, 16; see also p. 24 (discussing construction in terms of argument regarding input/output distinction)

Experts Agree on “Voice”

Transcript of Richard Martin Stern, Jr., Ph.D.
Conducted on October 30, 2025

1 UNITED STATES PATENT AND TRADEMARK OFFICE
2 BEFORE THE PATENT TRIAL AND APPEAL BOARD
3 ----- x
4 AMAZON.COM, INC., :
5 AMAZON.COM SERVICES LLC, :
6 and AMAZON WEB SERVICES, :
7 INC., :
8 Petitioners : IPR2025-00565
9 v. : U.S. Patent
10 SOUNDCLEAR TECHNOLOGIES LLC, : No. 11,069,337
11 Patent Owner. :
12 ----- x
13
14 Videotaped Deposition of
15 RICHARD MARTIN STERN, JR., Ph.D.
16 Conducted Remotely
17 Thursday, October 30, 2025
18 12:15 p.m. Eastern Standard Time
19
20 Job No. 604599
21 Pages: 1 - 119
22 Reported by: Janet A. Hamilton, RDR

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-2015, 54:1-6; POR, 7

Experts Agree on “Voice”

Transcript of Richard Martin Stern, Jr., Ph.D. Conducted on October 30, 2025

1	UNITED STATES PATENT AND TRADEMARK OFFICE
2	BEFORE THE PATENT TRIAL AND APPEAL BOARD
3	----- x
4	AMAZON.COM, INC., :
5	AMAZON.COM, INC., :
6	and 7
7	INC., :
8	:
9	v. SOUND CLEAR, INC.
10	SOUND CLEAR, INC.
11	- -
12	- -
13	- -
14	- -
15	- -
16	Conducted Remotely
17	Thursday, October 30, 2025
18	12:15 p.m. Eastern Standard Time
19	
20	Job No. 604599
21	Pages: 1 - 119
22	Reported by: Janet A. Hamilton, RDR

1 things like that. When I speak about speech

2 and/or voice, and I do use those terms somewhat

3 interchangeably, I'm talking about the subset of

4 things that humans utter that are directed toward

5 communication of an idea or a request or something

6 like that, normal conversation.

EX-2015, 32:1-6; POR, 7-8

12	Q So what's the plain and ordinary meaning
13	of the term voice that you apply when analyzing
14	the claims?
15	A The voice I consider to be the natural
16	productions that humans make in communicating with
17	each other via speech and also that humans make in
18	talking to machines using speech.

EX-2015, 28:12-18; POR, 7-8

Experts Agree on “Voice”

45. The word “voice” connotes the quality of information conveyed by a human. That information is not conveyed with simple changes in volume. It is conveyed via a multivariate notion of voice. The nature of a (singular) voice can change with volume. But volume alone does not create a new voice. Dr. Stem testified as follows:

48. A POSITA would have understood that the full context of the claim is helpful here. Voice is “classified.” It is a discrete concept. The incoming voice can either be one voice or one or more other voices. A POSITA would have understood that there is no middle ground. It is not a single variable that can have any value (on a continuous scale). It is not something that is understood to be calculated, measured,

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

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

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner.

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

DECLARATION OF JOSEPH MCALEXANDER

EX-2013, ¶¶45, 48; POR, 14

“Voice” Is Object of “Analyze”

Affect on “Analyze a Voice”

A. Construction reads out voice analysis

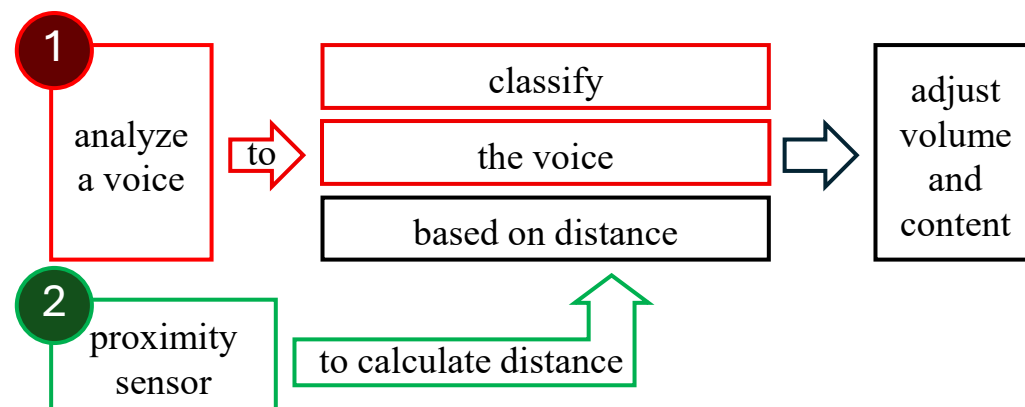
- the “voice” is analyzed
- ≠ distance being measured

B. Construction conflates two recited concepts:

1. analyze a voice, and
2. determine distance

POR, 7-17; POSR, 3-5

- **Limitation 1[b]:** a voice classifying unit configured ~~to analyze a voice spoken by a user and acquired by a voice acquiring unit~~ to classify [a] voice ~~spoken by a user and acquired by a voice acquiring unit~~ as either one of a first voice or a second voice based on the distance between the user and the voice-content control device;
- **Limitations 4[c] and 5[c]:** ~~analyzing the acquired voice to classify[ing] 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;

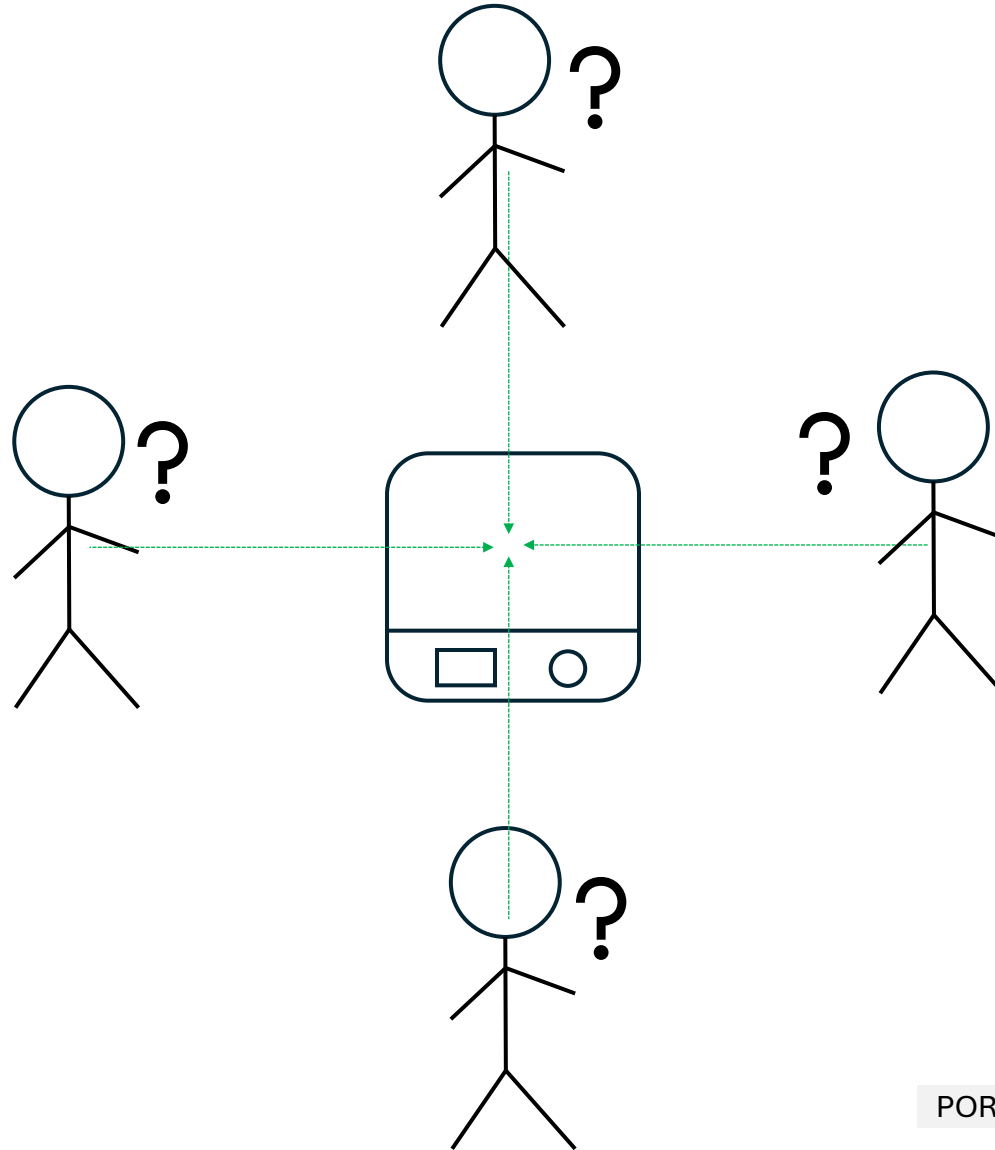


POR, 20; POSR, 2, 5-19

II.A

Prior Art at Most Classifies Distance

Impossible to Identify User or Voice Using Only Distance



POR, 25-26; POSR, 4-5, 9

Experts Agree on “Voice”

Transcript of Richard Martin Stern, Jr., Ph.D.
Conducted on October 30, 2025

1 UNITED STATES PATENT AND TRADEMARK OFFICE
2 BEFORE THE PATENT TRIAL AND APPEAL BOARD
3 ----- x
4 AMAZON.COM, INC., :
5 AMAZON.COM SERVICES LLC, :
6 and AMAZON WEB SERVICES, :
7 INC., :
8 Petitioners : IPR2025-00565
9 v. : U.S. Patent
10 SOUNDCLEAR TECHNOLOGIES LLC, : No. 11,069,337
11 Patent Owner. :
12 ----- x
13
14 Videotaped Deposition of
15 RICHARD MARTIN STERN, JR., Ph.D.
16 Conducted Remotely
17 Thursday, October 30, 2025
18 12:15 p.m. Eastern Standard Time
19
20 Job No. 604599
21 Pages: 1 - 119
22 Reported by: Janet A. Hamilton, RDR

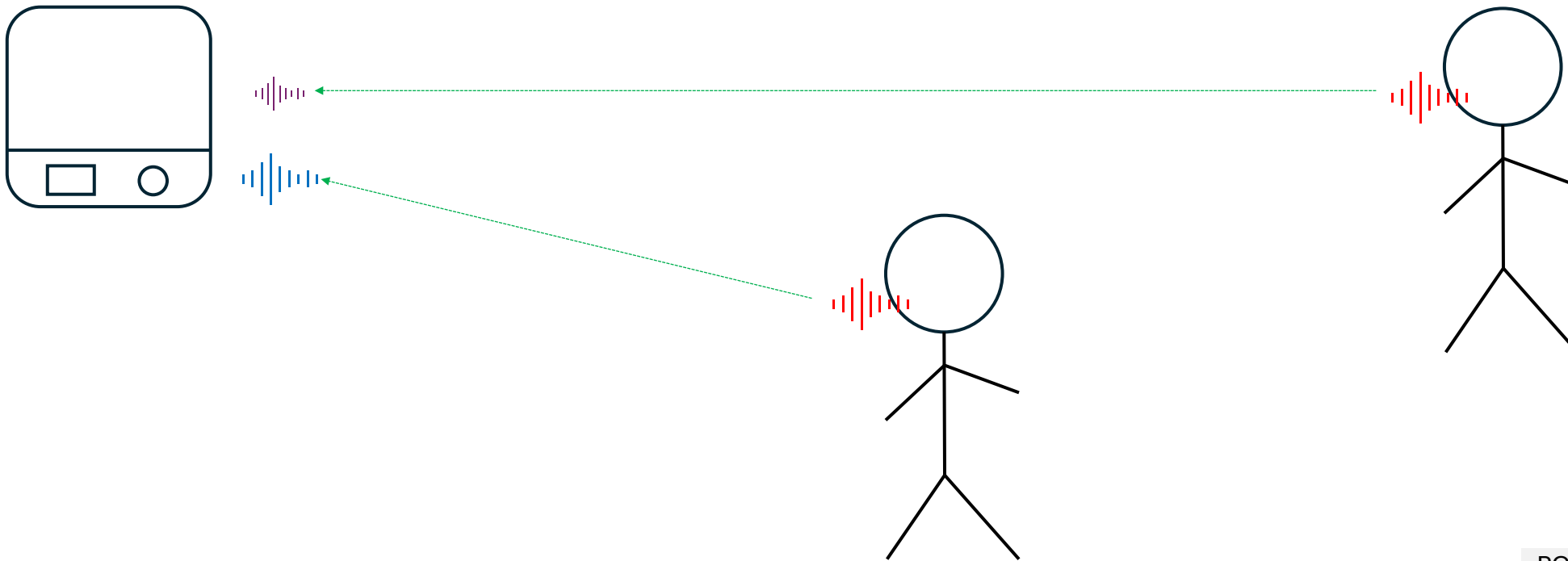
1	Q Does the distance between the user and the
2	device alone tell you anything about the
3	characteristics of the user's voice?
4	MR. REEVES: Objection. Form. Outside
5	the scope.
6	A In general, no.

EX-2015, 54:1-6; POR, 7

Perform Voice

Accordingly, for the purpose of our decision on institution, our preliminary claim construction only requires Petitioner to show a single user who may perform both a first and second voice depending upon the single user's changing distance (and associated changing voice volume) from the voice-content control device.

DI, 16

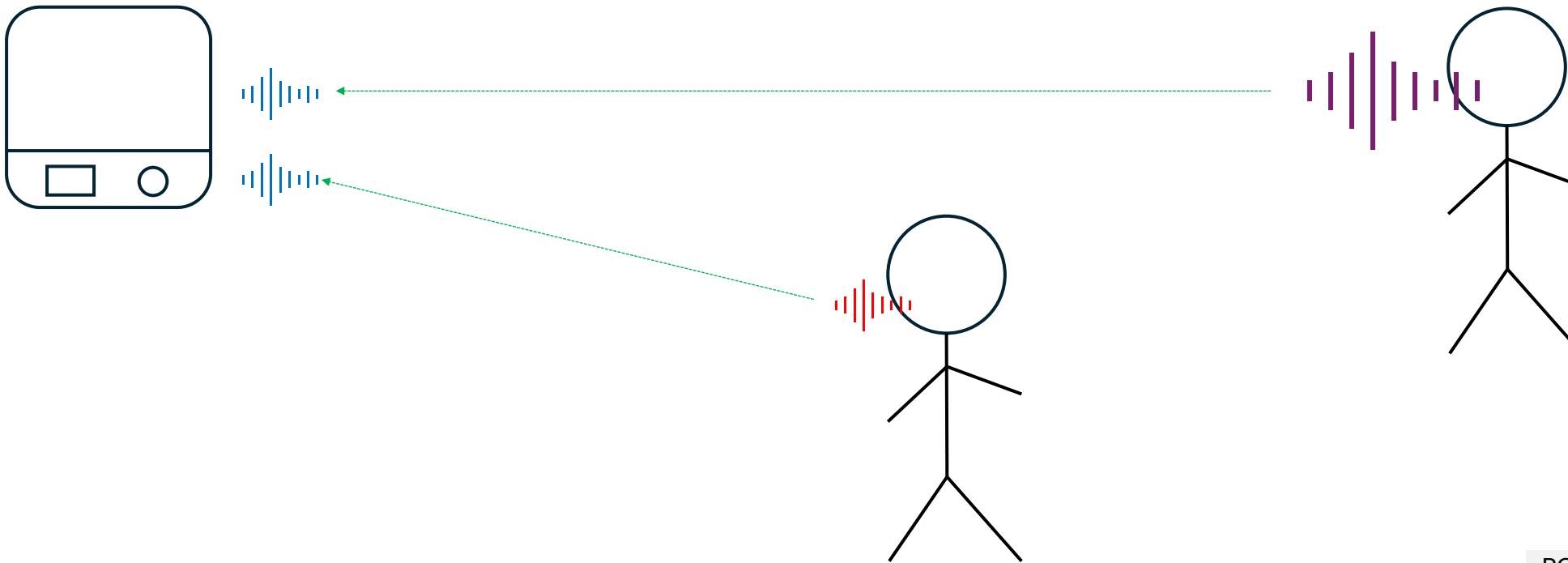


POR, 7-17, 25; POSR, 4-5, 9

Perform Voice

Accordingly, for the purpose of our decision on institution, our preliminary claim construction only requires Petitioner to show a single user who may perform both a first and second voice depending upon the single user's changing distance (and associated changing voice volume) from the voice-content control device.

DI, 16

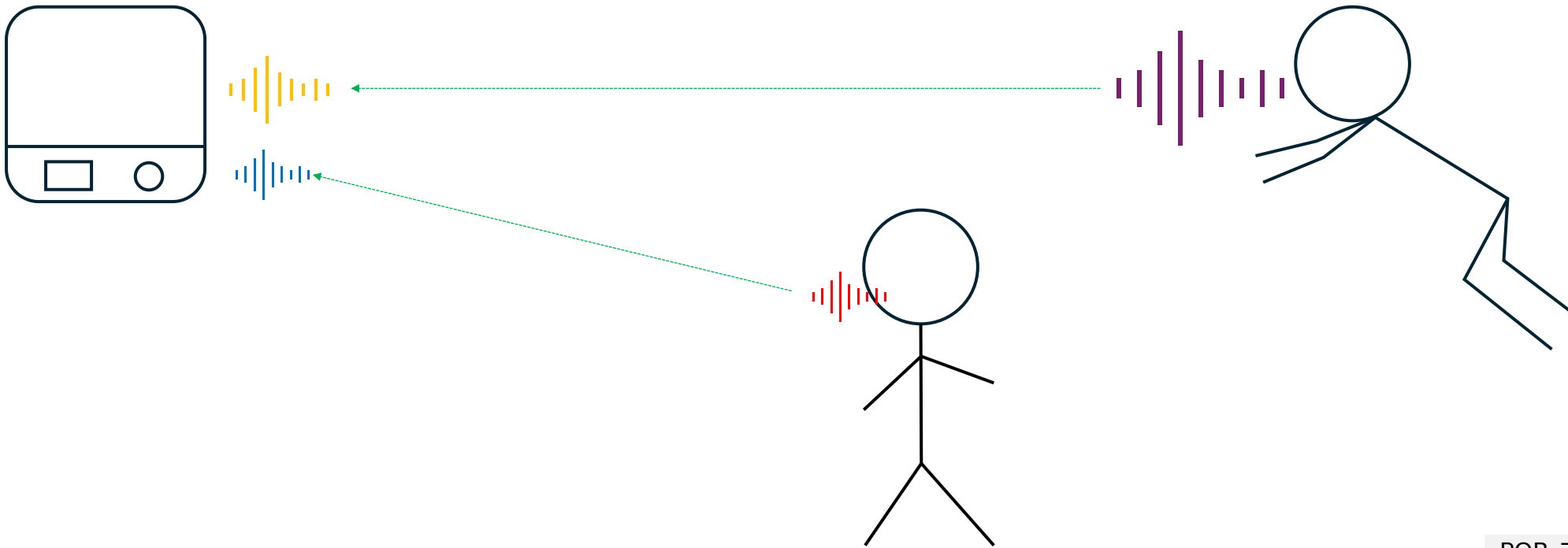


POR, 7-17, 25; POSR, 4-5, 9

Perform Voice

Accordingly, for the purpose of our decision on institution, our preliminary claim construction only requires Petitioner to show a single user who may perform both a first and second voice depending upon the single user's changing distance (and associated changing voice volume) from the voice-content control device.

DI, 16

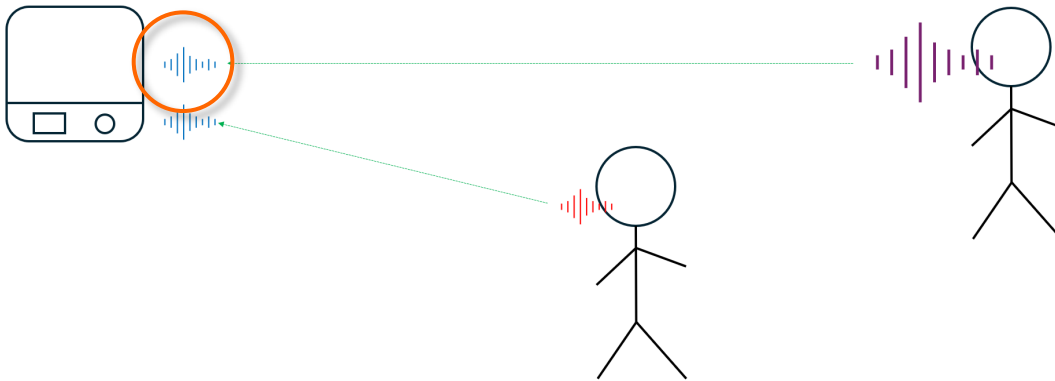


POR, 7-17, 25, POSR, 4-5, 9

Perform Voice

Accordingly, for the purpose of our decision on institution, our preliminary claim construction only requires Petitioner to show a single user who may **perform** both a first and second voice depending upon the single user's changing distance (and associated changing voice volume) from the voice-content control device.

DI, 16

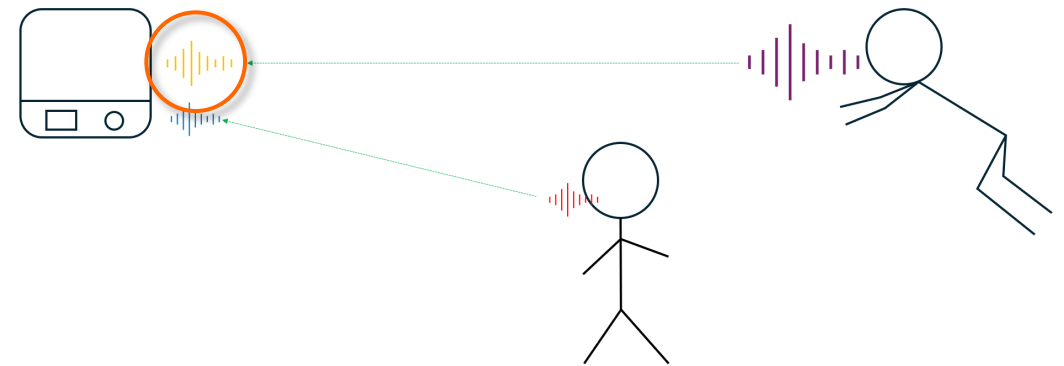


(57)

ABSTRACT

A **voice-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 classify the **voice** as either one of a first **voice** or a second **voice**, a process executing unit configured to analyze the acquired **voice** to execute processing required by the user, and a **voice-content**

EX-1001

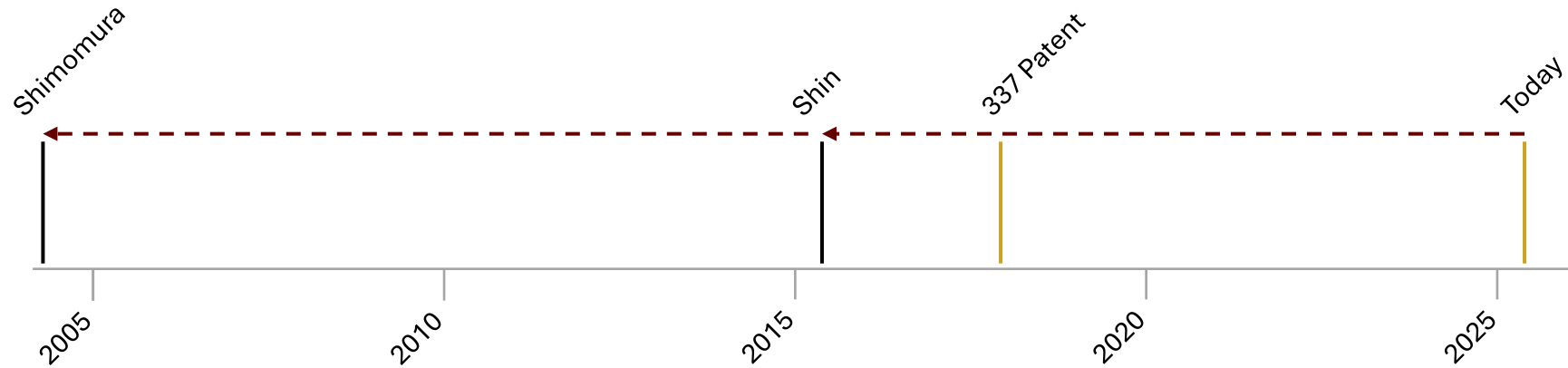


POR, 7-17, 25; POSR, 4-5, 9

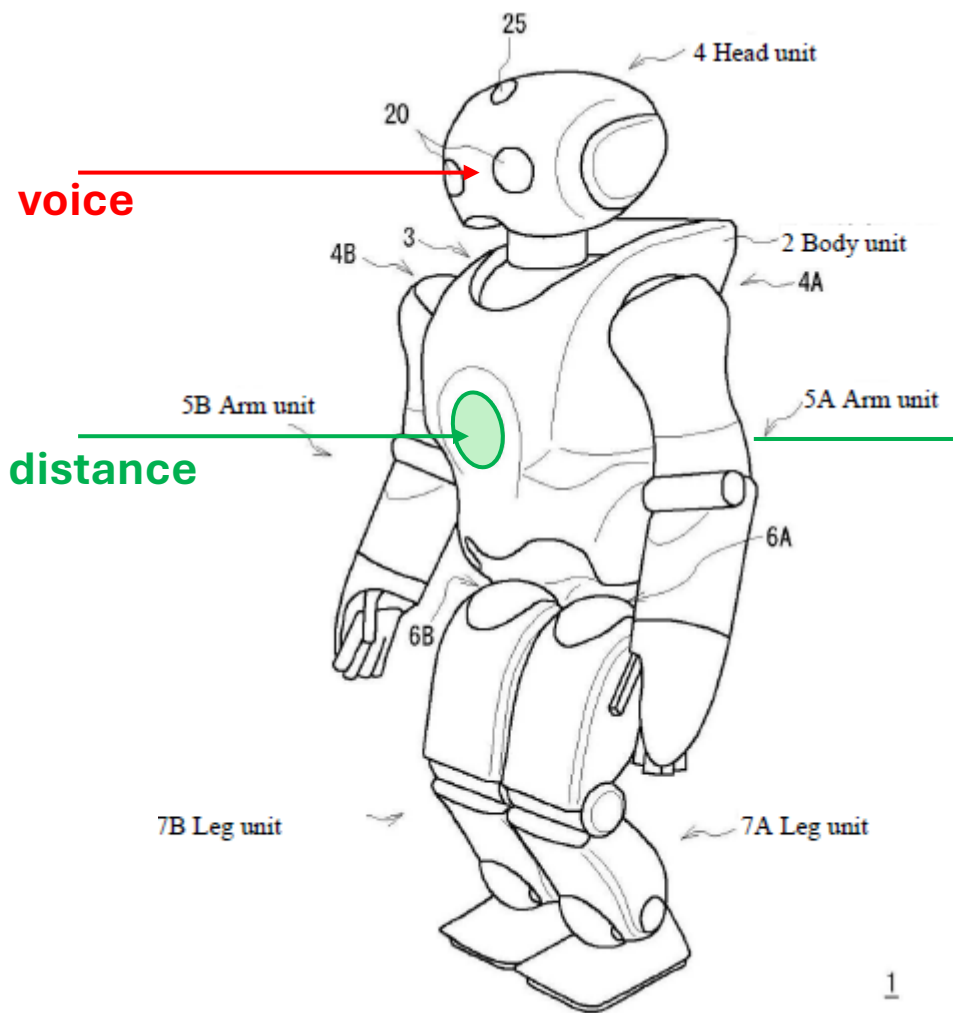
II.B

Prior Art Does Not Classify Voices

Decades of Hindsight



EX-1001, -1003, -1004



EX-1004 Fig. 1

change phrase

[FIG. 12]

Less than 200 cm	Less than 200 cm
None	Raise

55

FIG. 12 Phrase-end emphasis determination table

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)

volume

[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

TABLE 1

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

EX-1004 Figs. 9, 12

EX-1003 table 3 (altered and combined) and table 1

Just Distance

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE

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

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

TABLE 3	
Distance (D) between user and electronic device	Output amount of information of content
D < 1 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 (consolidated).)

Pet., 13 (citing Shin) (yellow annotation added)

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

[FIG. 7] Second Voice First Voice	
Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

Pet., 13, 39-40 (citing Shimomura) (yellow annotation added)

II.C

Input Versus Output

Black Box



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.

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

EX-2016, 62:22-63:8

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

6 MR. REYNOLDS: I apologize, Your Honor. I
7 will be very quick.
8 But the claim uses distance to classify the
9 speaker. It doesn't use distance directly to modulate
10 the volume. So you can't determine whether it meets
11 the claim just by knowing if one speaker gets closer to
12 the device or not and has the volume change because it
13 uses distance to figure out who's speaking. So you
14 would have to look at two different users and figure
15 out what they're speaking.

EX-1203 (Nov. 1, 2024 Hr'g), 68:6-15

II.D

Petition Lacks Support

“Classify”

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

Pet., 13

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

Pet., 39

63. Second, Shin discloses that the voice classifying unit (processor) analyzes the voice to classify it as either a first or second voice based on the distance between the user and the device. Shin’s processor executes a voice recognition application to process the voice 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 and its volume level, “based on the distance between the user and the electronic device 101[.]” (*Id.* ¶¶[0078], [0086]; *see also id.* ¶¶[0053], [0077]-[0080], claims 6, 17.) To do so, Shin’s processor classifies the voice based on the distance.

EX-1002, Stern Decl.

138. Second, Shimomura discloses that the voice classifying unit (control unit) analyzes the voice to classify it as either a first or second voice based on the distance between the user and the device. Shimomura’s control unit implements the behavior control system, which also includes a “speech synthesis unit” and a “speech character string transformation unit.” (Ex. 1004 (Shimomura) ¶¶[0036]-[0037], [0046], [0058], [0067], Fig. 4.) The speech synthesis unit and the speech character string transformation unit may “transform the character string” of the content spoken by the robot or change its “volume,” “speed,” “pause length,” or “intonation” “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, Stern Decl.

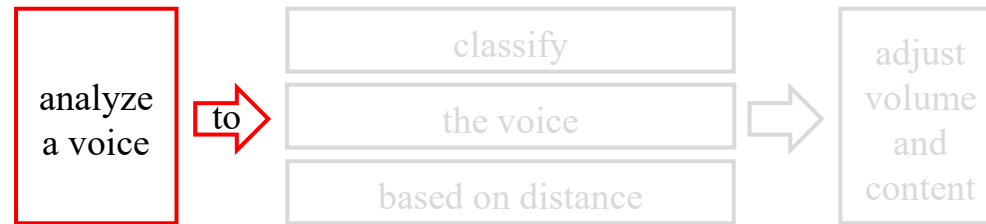
- not even attempt at inherency
- *Ipsissimis verbis* strawman for failure of inherency.
 - *Akzo N.V. v. United States Int'l Trade Comm'n*, 808 F.2d 1471, 1479 (Fed. Cir. 1986) (a reference disclosing “use of sulfuric acid” did not disclose “use of at least 98% concentrated sulfuric acid” because the 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, 10

III.

“Analyze a Voice”

Plain Language of Claims



EX-1001; POR, 12-13; POSR, 2, 5-19

Shin Analyzes Voice for Content Only—Not Classification

US 2017/0083281 A1

(19) United States
(12) Patent Application Publication
(31) Pub. No.: US 2017/0083281 A1
(43) Pub. Date: Mar. 23, 2017

(54) METHOD AND ELECTRONIC DEVICE FOR PROVIDING CONTENT

(71) Applicant: Samsung Electronics Co., Ltd., Gyeonggi-do (KR)
(72) Inventor: Sang Min SHIN, Gyeonggi-do (KR)
(73) Assignee: Samsung Electronics Co., Ltd.
(21) Appl. No.: 15/269,406
(22) Filed: Sep. 19, 2016
(30) Foreign Application Priority Data
Sep. 18, 2015 (KR) 10-2015-0132488

Publication Classification
(51) Int. Cl. G06F 3/16 (2006.01); G10L 25/63 (2006.01); G10L 25/78 (2006.01)
(52) U.S. Cl. G06F 3/163 (2013.01); G10L 25/78 (2013.01); G10L 25/63 (2013.01)
(57) ABSTRACT
An electronic device and a method are provided. The electronic device includes an audio input module configured to receive a speech of a user as a voice input, an audio output module configured to output content corresponding to the voice input, and a processor configured to determine an output scheme of the content based on at least one of a speech rate of the speech, a volume of the speech, and a keyword included in the speech, which is obtained from an analysis of the voice input.

START
OBTAIN SPEECH OF USER AS VOICE INPUT 502
DETERMINE RATE OF MOVEMENT OF USER THROUGH ANALYSIS OF VOICE INPUT 504
DETERMINE OUTPUT SCHEME OF CONTENT BASED ON DETERMINED RATE OF MOVEMENT OF USER 506
OUTPUT CONTENT CORRESPONDING TO VOICE INPUT BASED ON DETERMINED OUTPUT SCHEME 508
END

Amazon v. SoundClear
US Patent 11,069,337
Amazon Ex. 1003

-1-

US 2017/0083281 A1
Mar. 23, 2017

3

electronic device or may refer to a device (e.g., a device implementing an artificial intelligence) that uses an electronic device.

[0036] FIGS. 1A and 1B are diagrams illustrating an environment in which an electronic device operates, according to an embodiment of the present disclosure.

[0037] Referring to FIG. 1A, a user 10, who is spaced apart from an electronic device 100 by a specific distance and is at a standstill, may speak to the electronic device 100. The speech of the user 10 may be provided to the electronic device 100 as a voice input. For example, the user 10 may speak "Let me know what time it is now," with a moderate volume and at a moderate speech rate, and the speech may be provided to the electronic device 100 as a voice input. The electronic device 100 may receive the voice input through an embedded audio input module (e.g., a microphone) and may generate content corresponding to a result of analyzing the voice input. For example, the electronic device 100 may generate content, such as "The current time is nine ten AM", in response to a voice input, such as "Let me know what time it is now." The electronic device 100 may provide the generated content as sound through an embedded audio output module (e.g., a speaker). In this case, since the user 10 provides a voice input with a moderate volume and at a moderate speech rate at a standstill, the electronic device 100 may provide the content as sound with a moderate output volume level and at a moderate output speed. As such, the user 10 may be provided with the content corresponding to the voice input.

[0038] Referring to FIG. 1B, the user 10 who moves away from the electronic device 100 may speak to the electronic device 100. For example, the user 10 may be very busy with getting ready for work. Therefore, for example, the user 10 may make a speech, such as "What time is it now?", with a louder volume than usual and a faster speech rate than usual. The speech may be provided to the electronic device 100 as a voice input.

[0039] The electronic device 100 may receive the voice input through the embedded audio input module and may generate content corresponding to a result of analyzing the voice input. For example, the electronic device 100 may generate content, such as "nine ten", in response to the voice input, such as "What time is it now?". The electronic device 100 may provide the generated content as sound through the embedded audio output module. In this case, since the user 10 provides a voice input with a louder volume than usual and at a faster speech rate than usual while the user 10 moves away from the electronic device 100, the electronic device 100 may provide the content as sound with a relatively louder output volume level and a relatively faster output speed. As such, the user 10 may be provided with content corresponding to a voice input.

[0040] FIG. 2 is a diagram illustrating an electronic device according to an embodiment of the present disclosure.

[0041] Referring to FIG. 2, an electronic device according to an embodiment of the present disclosure may be implemented with the dedicated electronic device 100 that operates inside a house. The dedicated electronic device 100 may include various modules (e.g., elements of FIG. 3) for implementing embodiments according to the present disclosure, such as a driving system that is capable of providing the mobility to the electronic device 100 (e.g., a driving motor, various types of articulated joints for robots (e.g., a bipedal, quadrupedal robot), a wheel, a propeller, and the like), a camera that is capable of recognizing an input module that is capable of receiving a voice input, and the like.

[0042] Furthermore, the electronic device 100 may be implemented in a form in which a smartphone 201 and a docking station 202 are coupled to each other. For example, the smartphone 201 may provide a function for implementing the present disclosure through various modules (e.g., a processor, a camera, a sensor, and the like) embodied therein. Furthermore, for example, the docking station 202 may include a charging module (and power supplying terminal) that is capable of providing power to the smartphone 201, a driving system that is capable of providing the mobility (e.g., a driving motor, various types of articulated robotic joints, a wheel, a propeller, and the like), a high-power speaker, and the like.

[0043] A configuration of the electronic device, which is implementable in various ways as described above, will be described below with reference to FIG. 3. Elements to be described in FIG. 3 may be included, for example, in the electronic device 100 of FIG. 2 or in the smartphone 201 and/or the docking station 202. A content providing method of the electronic device 100 will be described with reference to FIGS. 4 to 8.

[0044] FIG. 3 is a block diagram illustrating an electronic device according to an embodiment of the present disclosure.

[0045] Referring to FIG. 3, an electronic device 101 according to an embodiment of the present disclosure includes a bus 110, a processor 120, a memory 130, an audio module 150, a display 160, a communication interface 170 and a distance detection module 180. The electronic device 101 may not include at least one of the above-described elements or may further include other elements (e.g., a camera, a sensor, and the like) in addition to the above-described elements.

[0046] For example, the bus 110 may interconnect the above-described elements 120 to 180 and may transfer (e.g., convey and/or data) among the above-described elements.

[0047] The processor 120 may include one or a central processing unit (CPU), an application processor (AP), or a communication processor (CP). For example, the processor 120 may perform an arithmetic operation processing associated with control and/or communication of the electronic device 101. For example, the processor 120 may execute a voice recognition application (e.g., S-Voice) to perform a content providing method according to an embodiment of the present disclosure.

[0048] According to embodiments of the present disclosure, the processor 120 may analyze a voice input received through an audio input module 151 and may output content corresponding to the voice input through an audio output module 152 in various schemes. For example, the content may be provided to a user as sound based on a text-to-speech (TTS) technology.

[0049] According to an embodiment of the present disclosure, the processor 120 may determine an output scheme of the content based on at least one of a speech rate of a user's speech, a volume of a user's speech, and a keyword included in the speech.

-15-

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

As discussed, Shin's processor analyzes the voice to identify the content the user is interested in, and then generates the requested content. (*Supra* §VII.A.4.)

Patent Owner

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

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

Pet., 25

[0048] According to embodiments of the present disclosure, the processor 120 may analyze a voice input received through an audio input module 151 and may output content corresponding to the voice input through an audio output module 152 in various schemes. For example, the content may be provided to a user as sound based on a text to speech (TTS) technology.

POR, 29-30

Petition Fails to Address Analysis of Voice

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT

AMAZON.COM
AMAZON.COM
AMAZON W
P

SOUNDCLEAR
Pat

Case No.
Patent

PETITION FOR
OF CLAIMS 1-5 OF

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

Pet., 13; POR 30-31

Shimomura Does Not Analyze Voice to Determine Distance or Classify

JP 2005-202076 A 2005.7.28

(19) Japan Patent Office (JP)

(12) Japanese Unexamined Patent Application Publication (A)

JP 2005-202076 A 2005.7.28

(51) Int. Cl. 7

G10L	13/00	F1	G10L	3/00
A63H	11/00		A63H	11/00
B25J	13/00		B25J	13/00
G10L	13/00		G10L	3/00

(71) Applicant: Japanese Patent Application 2004-7306 (P2004-7306)

(72) Inventor: January 14, 2004 (1.14.2004)

(74) Attorney: (72) Inventor

(54) [Title of the Invention] Speech control device and method and robot

(57) [Abstract]

[Problem]
Depending on the distance from the device or robot device to the conversation partner, the conversation partner may have difficulty hearing the speech of the device or robot device.

[Means for Solving]
By changing the speech style of the device or robot device when interacting with the user as necessary depending on the distance between the device or robot device and the user, the device or robot device can always speak in a style that is easy for the user to hear, allowing for smooth interaction with the user, thus realizing a speech control device and method, and a robot device that can improve entertainment value.

[Representative Drawing] Fig. 13

[0039]

The image recognition unit 41 performs image recognition processing and feature extraction, such as facial recognition and color recognition, based on the image signal S1 provided by the CCD camera 20. The image recognition unit 41 then sends out to the short-term memory unit 44 the various image recognition results, such as facial recognition information such as the person's unique facial ID (identifier), the position and size of the facial image region, and the color recognition information such as the position, size, and feature of the color region such as the color recognition result, and the image signal S1. The image recognition unit 41 also detects the distance to the imaging object by means of a so-called stereo visual method based on the image signal S1 from the CCD camera camera 20, and sends the detection results to the short-term memory unit 44.

[0059]

The behavior selection control unit 45 reads out character strings corresponding to the contents (words) to be spoken by the robot 1 when conversing with the user from the speech string database 47, and sequentially sends these as character string information to the speech character string transformation unit 51. At the same time, the behavior selection control unit 45 sends to the speech character string transformation unit 51 distance information indicating the distance to the conversation partner detected by the image recognition unit 41 and stored and held in the short-term memory unit 44.

EX-1004, ¶¶ 39, 59; POR, 43-44

FIG. 13 Speech change processing procedure

Amazon v. SoundClear
US Patent 11,069,337
Amazon Ex. 1004

Petition Fails to Address Analysis of Voice

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC.,
AMAZON.COM SERVICES, INC.,
AMAZON WEB SERVICES, INC.,
Petitioners,
v.
SOUNDCLEAR TECHNOLOGIES, INC.,
Patent Owner.

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

PETITION FOR INTER PARTIAL REEXAMINATION
OF CLAIMS 1-5 OF U.S. PATENT NO. 11,069,337

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

Pet., 39; POR 42-44

Comparison of Analyses

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

Pet., 13; POR 30-31

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

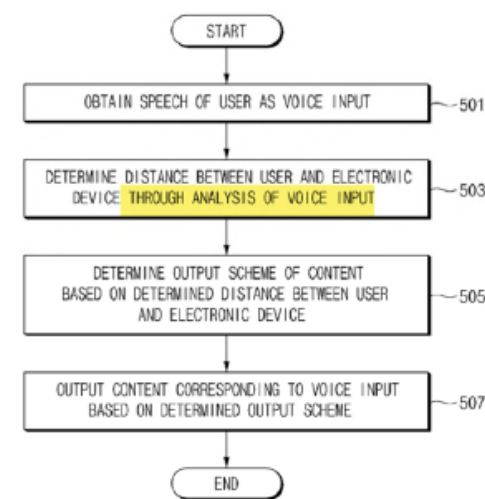
Pet., 39; POR 42-44

Petitioners' Last Resort

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

Pet., 13 (annotation added); *see also id.* 11-14

determining distance based on voice analysis, Shin discloses that. In Shin, user-to-device distance can be calculated by a proximity sensor, "by analyzing the voice input" and determining the distance "based on the obtained volume," or by using both methods together to determine distance more accurately. (EX-1003 ¶¶[0075]-[0077].) Again, Shin's Figure 5A clearly discloses that the distance may be determined "through analysis of voice input":



(*Id.*, Fig. 5A (annotated); EX-1202 ¶22.)

Reply, 15 (annotation in original)

Shin (Ground 1)

- Shin's voice analysis is only conducted to generate output content
- Attempt to add a theory too late
 - 37 C.F.R. § 42.23; *Rembrandt Diagnostics, LP v. Alere, Inc.*, 76 F.4th 1376, 1383 (Fed. Cir. 2023) (making “a meaningfully distinct contention” or relying on other embodiments for the first time are impermissible); *Yita LLC v. MacNeil IP LLC*, 69 F.4th 1356, 1367 (Fed. Cir. 2023) (affirming the Board's rejection of a “meaningfully distinct contention” presented for the first time in Petitioner's Reply).
- Attempting a two-for-one with no analysis

POR 31, n1; POSR, 11-15

Shin (Ground 1)

- the claimed distance calculation and voice analysis are separate limitations
- Table 3 of Shin does not have voice as an input or output (text content)

POSR, 14-15

1[pre]. 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;

POR 31, n1; POSR, 11-15

IV.

Obviousness



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.

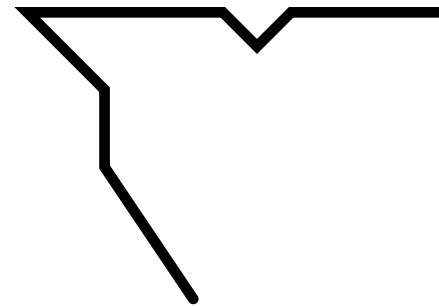
IPR2023-00701 *Artic Cat Inv. v. Bombardier Rec. Prods.*, 876 F.3d 1350, 1358; Sur-Reply, 3-4

POR, 55-65; Sur-Reply, 18-22



Dr. Stern's testimony does not alter this analysis. PO argues that Dr. Stern was "unable to articulate the *differences* between [the prior art] and claim 1 of the '337 patent beyond limiting the claim to a single threshold value." (POR, 57.) But, as the Board recognized, "no modifications are required to arrive at the claimed invention" under Ground 2 (and the same reasoning applies to at least Ground 3). (DI, 18.) Thus, there are no differences to articulate. PO's suggestion that there must be *some* difference to find obviousness contravenes the bedrock principle that "anticipation is the epitome of obviousness." (*Id.* (citing *Jones*, 727 F.2d at 1529).)

Reply, 57; POR, 57; Sur-Reply, 18-22



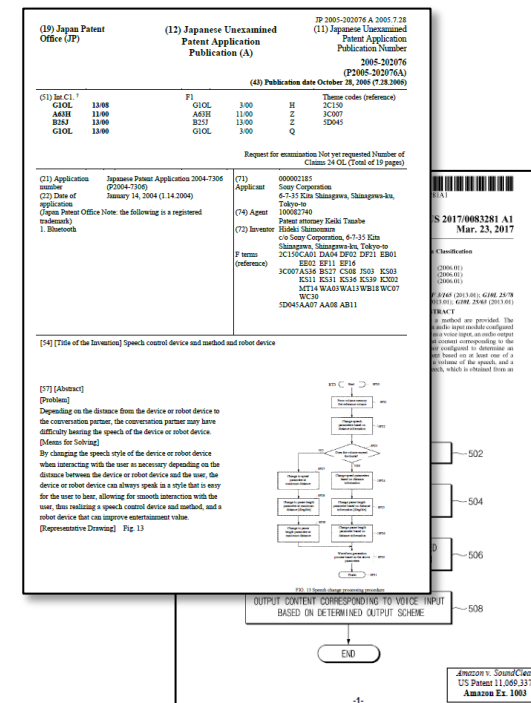
Ground 3

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

Pet. 47

- **Shimomura:** teaches transforming output content when further increase in volume is not possible. Thus, a POSITA would have understood from Shimomura that it is unnecessary to transform output content when increase in output volume is sufficient for the robot’s speech to be understandable by the user.
- No motivation to make the distances equal. it would undermine the purpose of Shimomura

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



POR, 55-65; Sur-Reply, 18-22

[A] modification such as the one shown in the first modified table would have been the simple substitution of a known element ([disclosure]) for another known element ([disclosure]) to obtain predictable results (a device that [contains both disclosures]). (*Id.* ¶164)

Pet. 48

B. Simple Substitution of One Known Element for Another To Obtain Predictable Results

To reject a claim based on this rationale, Office personnel must resolve the *Graham* factual inquiries. Then, Office personnel must articulate the following:

- (1) a finding that the prior art contained a device (method, product, etc.) which differed from the claimed device by the substitution of some components (step, element, etc.) with other components;
- (2) a finding that the substituted components and their functions were known in the art;
- (3) a finding that one of ordinary skill in the art could have substituted one known element for another, and the results of the substitution would have been predictable; and
- (4) whatever additional findings based on the *Graham* factual inquiries may be necessary, in view of the facts of the case under consideration, to explain a conclusion of obviousness.

MPEP § 2143

[A] modification such as the one shown in the first modified table would have been the simple substitution of a known element ([disclosure]) for another known element ([disclosure]) to obtain predictable results (a device that [contains both disclosures]). (*Id.* ¶164)

Pet. 48

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.

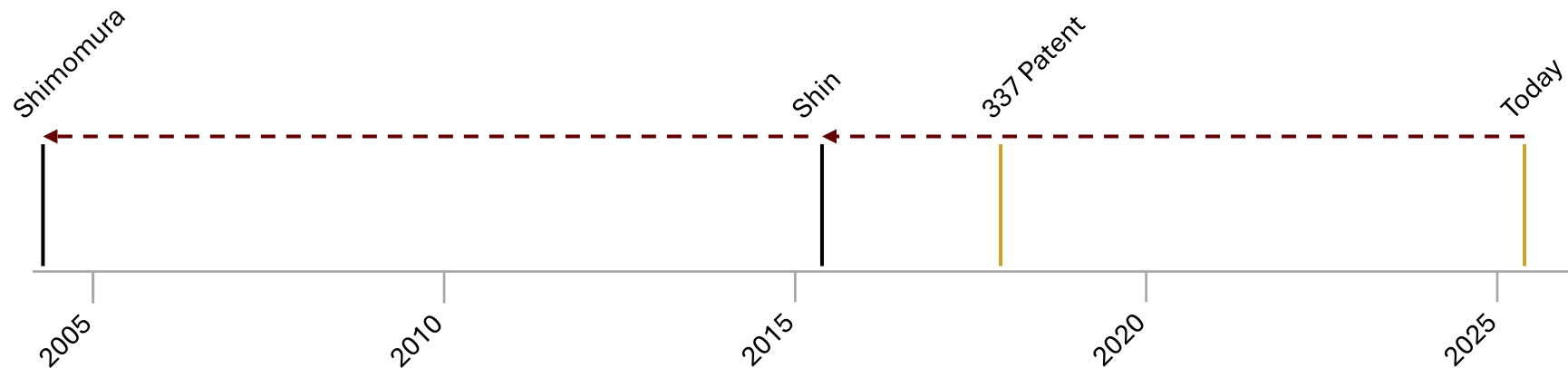
Just as it was possible to begin with the objective to upgrade Asano to work with a computer-controlled throttle, so too was it possible to take an adjustable electronic pedal like Rixon and seek an improvement that would avoid the wire-chafing problem. Following similar steps to those just explained, 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, 550 U.S. at 407-09 and 424-26

Illogical Combination

- Shin and Shimomura chose simplistic measurements
- Both set specific distances for specific reasons
- Both received “voices”
- Both declined to classify “voices”
- Shimomura specifically decided to disarticulate *distance* and *volume*
- No obviousness grounds identifies how or why *any* alleged modification would have improved the device



EX-1001, -1003, -1004; POR 55-64

Shimomura

- Chose the thresholds included in the tables for a reason.
- The Petition's obviousness positions destroy the
- teaches transforming output content when further increase in volume is not possible. Thus, a POSITA would have understood from Shimomura that it is unnecessary to transform output content when increase in output volume is sufficient for the robot's speech to be understandable by the user.
- No motivation to make the distances equal. it would undermine the purpose of Shimomura

POR 47-50; POSR, 20-21

Reply

- No response to these points
- Attempts to shift the burden to Patent Owner
- Patent Powner: changing the tables would undermine Shimomura
- Petitioner: That does not mean a POSITA would not have been motivated to change them
- Record: blank on motivation

Reply, 20-22