

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

PETITIONERS' REPLY TO PATENT OWNER'S RESPONSE

TABLE OF CONTENTS

I.	CLAIM CONSTRUCTION -----	1
A.	“User”-----	1
B.	“Classify”-----	1
C.	“Voice” -----	5
D.	Analyze a Voice to Classify it Based on Distance -----	7
II.	GROUND 1 AND 2: SHIN -----	9
A.	Shin Discloses Classifying Voice Based on Distance -----	9
B.	Shin Discloses Voice Analysis. -----	14
C.	PO’s Remaining Arguments Fail-----	15
D.	Shin Renders the Claims Obvious -----	16
III.	GROUND 3: SHIMOMURA-----	17
A.	Shimomura Discloses Voice Classification-----	18
B.	Shimomura Discloses Voice Analysis-----	19
C.	Shimomura Renders the Claims Obvious -----	20
1.	Limitation 1[h]-----	20
2.	Limitations 1[f]-1[g] -----	22
D.	The Shimomura Translation Is Accurate-----	23
IV.	GROUND 4: SHIMOMURA AND SHIN-----	23
V.	PO’S ADDITIONAL ARGUMENTS ARE UNAVAILING -----	25
A.	A POSITA Would Have Reasonably Expected Success -----	25
B.	The Petition Applies the <i>Graham</i> Factors-----	26
C.	PO’s Remaining Arguments Fail-----	26
VI.	CONCLUSION -----	27

TABLE OF AUTHORITIES

<i>Cases:</i>	<i>Page(s):</i>
<i>DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.</i> , 567 F.3d 1314 (Fed. Cir. 2009)-----	21
<i>Global Shade Corp. v. With-U E-Commerce (Shanghai) Co., Ltd.</i> , IPR2023-00580, Paper 34 (P.T.A.B. Aug. 7, 2024) -----	21-22
<i>Hum Indus. Tech., Inc. v. Amsted Rail Co., Inc.</i> , IPR2023-00540, Paper 10 (P.T.A.B. Aug. 11, 2023)-----	27
<i>In re May</i> , 574 F.2d 1082 (C.C.P.A. 1978)-----	10, 18
<i>Jones v. Hardy</i> , 727 F.2d 1524 (Fed. Cir. 1984)-----	17
<i>Luxottica of Am. Inc. v. E-Vision Optics, LLC</i> , IPR2024-01069, Paper 33 (P.T.A.B. Dec. 16, 2025) -----	25-26
<i>Merck & Co. v. Biocraft Labs., Inc.</i> , 874 F.2d 804 (Fed. Cir. 1989) -----	17
<i>ResMed Corp. v. Cleveland Med. Devices, Inc.</i> , IPR2025-00159, Paper 10 (P.T.A.B. June 12, 2025) -----	25
<i>UCB, Inc. v. Actavis Labs. UT, Inc.</i> , 65 F.4th 679 (Fed. Cir. 2023) -----	21
<i>Whitserve, LLC v. Comput. Packages, Inc.</i> , 694 F.3d 10 (Fed. Cir. 2012)-----	10, 18
<i>Xilinx, Inc. v. Godo Kaisha IP Bridge 1</i> , IPR2017-00842, Paper 36 (P.T.A.B. June 7, 2018)-----	23-24

The prior art plainly discloses exactly what the '337 patent claims and describes. Faced with that art, Patent Owner (“PO”) shifts theories and attempts to distort the claims through strained interpretations devoid of support in the claim language or specification. The Board should reject those arguments.

I. CLAIM CONSTRUCTION¹

A. “User”

The Board construed “user” as “one or more users.” (Paper 12 (“DI”), 15.) That construction rejected PO’s argument that the claims required voices from *different* users. (Paper 9 (“POPR”), 23.) PO and its expert now admit the Board’s construction is correct, but PO argues that “no construction of ‘user’ is necessary.” (Paper 23 (“POR”), 5-6; EX-1201, 30:7-23.) That is wrong. The Board’s Final Written Decision should adopt its preliminary construction, which was necessary to resolve a disputed issue.

B. “Classify”

Abandoning its “different users” theory, PO now asserts that “classify” requires a classification based on something *other than* the user-to-device distance. (POR, 6, 19-20.) This argument contradicts PO’s prior representations, PO’s expert, the claim language, and the specification.

¹ PO’s assertion that Petitioners will seek different constructions in district court (POR, 21) is unfounded.

In district court, PO repeatedly represented that voice is classified based on distance alone. (EX-1203, 45:4-10 (PO stating that “[c]lassifying the incoming voice as a first recorded voice or a second recorded voice, that’s a binary decision ... [and] we did it with the distance calculation”), 68:6-9 (“the claim uses distance to classify the speaker”), 54:6-7 (describing invention as “a voice-controlled device that operates *using distance to classify voices*”) (emphasis added); *see also* EX-1204, 10 (arguing claim 4 “recites calculating a specific distance, acquiring voice spoken by a user, and classifying the voice ... based on the distance”).)

PO’s expert’s testimony confirms that the voice can be classified based on distance alone. He explained that, *in one embodiment*, the voice classifying unit can “determine if a voice is a whisper or not and classify the voice accordingly.” (EX-2013 ¶37.) But “[o]ther embodiments of the voice classifying unit use the ... distance between the speaker and the device[.]” (*Id.* ¶38 (emphasis added).) He admits that the specification describes comparing the “user-to-device distance” and a threshold “for the purpose of distinguishing one class of voice from another.” (*Id.* ¶52.) He admits that the “claimed classifying unit uses those characteristics to classify voices” and that each characteristic—including distance—“can be used *alternatively* or in different combinations.” (*Id.* ¶53 (emphasis added).) Indeed, he concedes that a POSITA may use only “one variable” to classify the voice, even if it is not “perfectly effective.” (*Id.* ¶54.) Thus, PO’s expert confirmed that the voice

classifying unit may classify the voice based solely on the user-to-device distance.

Petitioners’ expert agrees. (EX-1202 ¶5.)

The ’337 patent also confirms that the classification can be based solely on distance. The claims explicitly recite voice classification “based on the distance”—and nothing else—and the specification explicitly discloses the same. (EX-1001, claim 1 (“classify the voice ... *based on the distance* between the user and the voice-content control device” (emphasis added)), 8:20-26 (“the distance can be used as a feature value to perform the classification to the first voice V1A and the second voice V1B”), 8:30-34 (“the voice classifying unit 38 sets a threshold of the feature value, and classifies the voice V1 ... based on whether the feature value exceeds the threshold”), 8:10-12; EX-1202 ¶6.) Thus, the claims, specification, both experts, and PO’s representations to the district court all confirm that the claimed classification may be based on distance alone.

PO’s arguments to the contrary are meritless. PO first argues that distance-only classification would “convert the recited ‘classification’ into a measurement.” (POR, 6.) That is incorrect. The claims recite, and the specification describes, that a single user’s voice may be classified into one of two categories—a “first voice” or “second voice”—based on a distance measurement.

PO next argues that “classify” “require[s] something discrete, not a mere measurement” of “a single, continuously changing variable (like volume or

distance).” (POR, 6.) This makes no sense. Measurements are discrete. The specification describes classifying the voice based on measurable characteristics such as distance, volume, frequency, etc. (EX-1001, 8:20-34.) And, the claims expressly require classifying based on a measurable characteristic—distance. (EX-1202 ¶7.)

PO next argues the claims require classifying between “at least two separately identifiable voices” because the claims allegedly recite “two (or more) distinct ‘voices’: a first and second voice.” (POR, 6.) This is wrong. Both experts agree the claims recite receiving “a voice” from “a user”—which includes a *single* voice from a *single* user—and placing it into a separately identifiable *class* (i.e., category). (EX-1202 ¶8; EX-1201, 30:7-23, 61:24-63:8; EX-1205, 120.) Regardless, in a device with a proximity sensor, a voice received from one distance is “separately identifiable” from a voice received from another distance. PO’s expert admitted this. (EX-1201, 64:6-15 (distance is a “class[] of separately identifiable voices”).)

PO does not propose any construction for “classify” and Petitioners do not believe any construction is necessary, as it should be given its plain and ordinary meaning. However, if the Board were to construe this term, it should be construed to mean “arrange in categories,” which is its plain and ordinary meaning. (EX-1202 ¶9; EX-1205, 120; EX-1201, 29:20-30:23.)

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).

PO’s arguments regarding “volume” are irrelevant because none of Petitioners’ invalidity arguments depend on volume classifications. Nevertheless, Petitioners address PO’s incorrect arguments here for completeness.

PO’s assertion that “[v]oice’ does not equate to [a] mere incidental change in volume” and that a POSITA would have had “a multivariate notion of voice” misses the point. (POR, 7-17.) The issue is not whether a voice has multiple variables or characteristics, but which variable or characteristic is used to classify the voice. The ’337 patent discloses, and both experts agree, that a voice may be classified based on a single variable, such as the volume *or* distance. (EX-1001, 7:64-8:34; EX-1202

¶10; EX-1201, 64:6-15.)

The testimony PO cites (POR, 8-9) is irrelevant. Dr. Stern simply explained that a voice has many characteristics, and that increasing volume is usually associated with other changes that could alternatively be used for classification. (*Id.*) This is consistent with the '337 patent. (EX-1202 ¶11.)

PO argues that the Board's construction would mean that "every 'chang[e in] voice volume' would be deemed to be a new voice[.]" (POR, 9.) Not so. The specification is clear that the system may set "a threshold of the feature value" (e.g., distance² or volume), and may classify the voice "based on whether the feature value exceeds the threshold." (EX-1001, 8:30-34.) Thus, for example, all changes below the threshold would not result in a different classification. (EX-1202 ¶13.)

PO next argues that the Board's construction conflates voice, volume, and distance. (POR, 10-17.) It does not. PO's argument to the contrary appears to be based on a misunderstanding of the Board's institution decision. (*Id.*)

For the reasons above, the Board should construe "a user" to mean "one or more users." No separate construction for "voice" or "classify" is necessary to resolve the invalidity issues here, and PO has offered none.

² PO later argues that Dr. Stern "admits that while distance may sometimes be sufficient to classify voice, it is not always sufficient." (POR, 14.) This is a mischaracterization; Dr. Stern actually testified that distance is not always sufficient to detect whether a user is whispering. (EX-2015, 48:19-49:7.)

D. Analyze a Voice to Classify it Based on Distance

PO argues that the claims “require **(1)** a classification that must be based on **(2)** an analysis of the acquired voice and **(3)** a distance[.]” (POR, 17-21.) But there is no need to rewrite the claims, which recite a device configured to “analyze a voice spoken by a user ... to classify the voice as either one of a first voice or a second voice based on the distance between the user” and the device. (EX-1001, claim 1.)

PO is correct that the distance *calculation* is not the claimed analysis. The proximity sensor calculates distance. The voice classifying unit then analyzes the voice by determining whether its distance exceeds a threshold. (EX-1001, 8:20-34.) That analysis allows the classifying unit to “classify” the voice as being nearer (“first voice”) or farther (“second voice”) than the threshold. (*Id.*) Thus, the analysis includes comparing the voice’s distance to the threshold, and the result is classification as a “first voice” or “second voice.” (EX-1202 ¶15.) Again, this is consistent with PO’s characterization of its claims in district court and with PO’s expert’s testimony. (EX-1203, 45:4-10, 68:6-9, 54:6-7; EX-1204 at 10; EX-1201, 70:2-17 (distance is an input to the analysis necessary to “conclude the classification”), 67:12-24 (the analysis uses an input “to resolve in a unique classification”).)

PO next argues that the *specification* “consistently describes” analyzing a voice to determine whether it is a whisper and classifying it on that basis. (POR, 18-

19.) But it is undisputed that the specification discloses whisper as merely one potential basis for classification and that “other embodiments” classify voice based on other variables, including distance. (EX-1001, 7:51-8:34; EX-2013 ¶¶38; EX-1202 ¶¶12, 16 (“a POSITA would have understood the specification to disclose that voice can be classified ‘by using any method,’ including distance alone”); EX-2015, 36:7-37:6.) Indeed, both experts agree that the claims do not require any whisper-based classification. (*Id.*; *see also* EX-1201, 35:10-14, 54:21-55:3.) Thus, PO’s suggestion that the claims require classification based on something other than distance, such as whether the received voice is a whisper, is wrong.

PO argues that classification cannot be based on distance alone for two reasons. (POR, 19-20.) First, PO argues that such a construction reads limitations out of the claim. Not so. As discussed above, the classification is the output of the analysis (e.g., comparison to threshold) that relies on the calculated distance. This is consistent with the patent and PO’s own explanation of its claims.

Second, PO argues that the Board’s interpretation conflates claim limitations because *determining* the distance is not the claimed analysis. But the Board never suggested that the distance calculation *is* the analysis. The Board merely stated that the claim only requires a single user and that that user’s voice may be classified based on the user’s changing distance. That is entirely consistent with the claim language, the specification, and PO’s description of its claims in district court.

Thus, the Board should reject PO’s attempts to implicitly construe the claim to require further “analysis” or “classification.”

II. GROUNDS 1 AND 2: SHIN

A. Shin Discloses Classifying Voice Based on Distance.

PO first argues that Petitioners failed to show that Shin discloses “basing *any* function or output on ‘voice.’” (POR, 23-24.) PO also claims that the only express disclosure of “voice” in Shin relates to generating content. (*Id.*, 24.) PO is mistaken.³

Shin expressly discloses “analyz[ing] a voice input[.]” (EX-1003 ¶[0048].) Indeed, Shin discloses “receiv[ing] a speech of a user *as a voice input* ... to determine an output scheme ... from *an analysis of the voice input*[.]” (*Id.*, Abstract (emphases added).) These disclosures are expressly cited in the Petition. (*See, e.g.*, Pet., 12 (citing EX-1003 ¶[0048], Abstract, ¶¶[0003]-[0011], [0037]-[0039], [0041], [0060]-[0061], [0155], [0158], claims 1, 15, 20).) Indeed, Shin discloses analyzing a voice input—for example, evaluating speech rate, volume, and/or user-to-device distance—to alter output, just like the ’337 patent. (*See, e.g.*, EX-1003 ¶¶[0049], [0053].) PO cannot reasonably argue that Shin fails to disclose analyzing a voice, as Shin’s Figures also plainly disclose this limitation:

³ PO also dismisses the Petition as “conclusory,” quoting the topic sentence of a paragraph while ignoring the supporting sentences after it. (POR, 24 (discussing Pet., 13).)

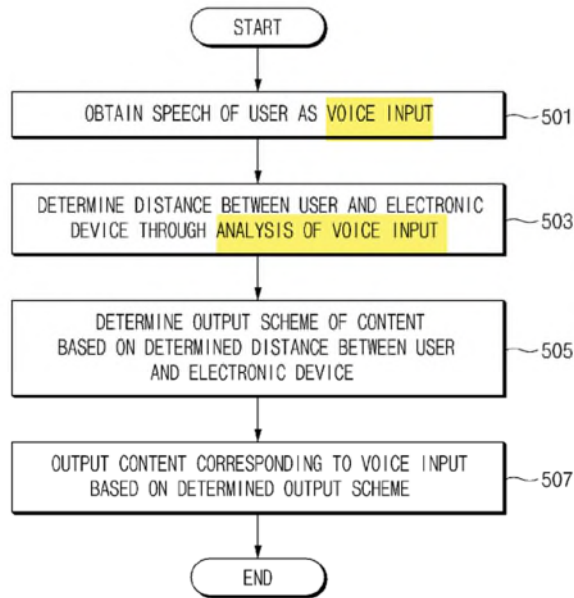


FIG. 5A

(*Id.*, Fig. 5A (annotated); EX-1202 ¶17.)

PO next argues that Shin lacks express disclosure of classification. (POR, 24.) PO’s argument is based on the mistaken premise that the reference must use those exact words. *In re May*, 574 F.2d 1082, 1090 (C.C.P.A. 1978) (reference need not describe claimed subject matter in *ipsissimis verbis*); *Whitserve, LLC v. Comput. Packages, Inc.*, 694 F.3d 10, 21 (Fed. Cir. 2012). The Petition and Dr. Stern each explained how Shin expressly discloses voice classification. (Pet., 11-14 (citing EX-1002 ¶¶59-66).)

PO next argues that Shin’s Table 3 discloses “only a calculation of distance and nothing else.” (*Id.*) But Table 3 does not disclose a distance *calculation* at all.

Shin discloses that the proximity sensor and, optionally, the processor calculate distance. (See, e.g., EX-1003 ¶¶[0075]-[0077], claim 9.)

Instead, Table 3 expressly discloses *classifications based on distance*. (Pet., 11-14; EX-1202 ¶18.)

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

(EX-1003, Table 3 (consolidated).) The left column includes categories (or classifications) for the distance D. (*Id.*) And the right column describes the output resulting from each such classification. (*Id.*)

This is exactly how the '337 patent describes voice classification. It describes two classes: (1) $D \leq x$; and (2) $D > x$, where x is a threshold value. (EX-1001, 8:30-34; EX-1202 ¶19.) The device receives a voice, calculates the user-to-device distance D, determines in which distance-based category (class) the received voice belongs, and provides appropriate output. (*Id.*; EX-1001, 8:60-66.) Shin discloses the same thing. The only difference is that the '337 patent gives the two classes generic names: “V1A” (or “first voice”) and “V1B” (or “second voice”). Shin does

not expressly refer to the classes’ *names*, but still discloses the classifications.⁴

PO’s irrelevant hypothetical argues that Shin would be incapable of distinguishing between “two users, with two different voices, [who] were standing [at] the same distance.” (POR, 25.) But PO now admits the claims do not require distinguishing between two different users’ voices. When two users are the same distance from the device in the ’337 patent (e.g., exceed the distance threshold), each user’s voice would fall into the second class ($D > \text{threshold}$), both would be classified as “V1B” (or “second voice”), and the output in response would include the same content at the same volume. The same is true for Shin.

PO next argues that “a POSITA would never have understood Shin’s disclosure of distance to be synonymous with ‘voice’” or with “‘a classification.’” (POR, 26.) But that is not what Petitioners argued. Rather, as discussed, Shin’s Table 3 discloses more than just distance determination—it discloses voice classifications based on distance. (EX-1202 ¶18.)

Dr. Stern’s testimony does not support PO. For example, PO’s assertion that Dr. Stern testified that Shin’s Table 3 “says ‘nothing’ other than ‘about the distance between the user and the electronic device’” is misleading. (POR, 27-28.) Dr. Stern testified that, in the context of the ’337 patent, voice classification depends on the

⁴ For the same reason, PO’s argument that “Shin directly correlates content output to distance” and thus “does not classify the voice input in any way” is incorrect. (POR, 26; EX-1202 ¶20.)

“input to the device.” (EX-2015, 24:9-17.) He also explained how Shin’s Table 3 discloses voice classification. (*Id.*, 71:18-72:7 (“you can have voice classification without the explicit use of the term voice, and here the classification is based on distance as taught by the specifications [and] the claims of the patent”), 68:16-69:18, 70:16-71:6.) For the specific questions quoted in the POR, PO’s counsel made it clear that he was *not* asking about classification. (EX-2015, 72:20-22 (“[Y]ou keep answering in the context of classification. That’s not my question, sir.”).) PO cannot now argue that Dr. Stern’s testimony was an admission that Table 3 does not disclose voice classification.

Finally, PO’s argument regarding a system that always classifies a voice as different “simply because distance is changed” (POR, 28-29) is not relevant because that is not what Shin discloses. Only certain distance thresholds result in different voice classifications:

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.) That is the same as the ’337 patent.

Thus, Shin plainly discloses classifying a user’s voice into one of two classes

(e.g., “first voice” or a “second voice”) based on the user-to-device distance.

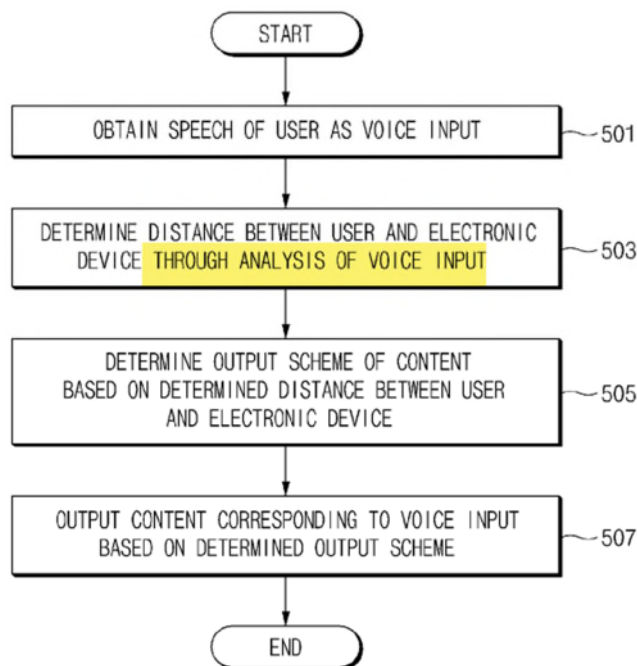
B. Shin Discloses Voice Analysis.

Contrary to PO’s assertion (POR, 29-32), Shin discloses analyzing the acquired voice to classify it based on distance. Shin explicitly discloses “analyzing” the voice input. (*E.g.*, EX-1003 ¶¶[0037], [0048], [0076]-[0077] (user-to-device distance may be determined by “analyzing the voice”), Abstract.) Shin also explicitly discloses analyzing the voice by comparing user-to-device distance to threshold values to classify the voice into one of several classes, so the device can generate appropriate output (e.g., content, volume). (*Id.* ¶¶[0076]-[0080].) This is the same “analysis” as in the ’337 patent, and the claims require no more. (EX-1202 ¶21.)

PO attempts to dismiss Shin’s ¶[0048] by arguing that it relates only to the “content” of speech. (POR, 30.) But it is not so limited. More importantly, PO ignores the substantial additional disclosure contained in Shin. (*See* Pet., 12-13 (citing EX-1003, Abstract, ¶¶[0003]-[0011], [0037]-[0039], [0041], [0047], [0051], [0060]-[0061], [0062], [0077]-[0081], [0086]-[0087], claims 1, 6, 15, 17, 20).)

PO argues that Shin does not “determine[] the user-to-device distance ... based on any voice analysis[.]” (POR, 31.) PO misses the point. First, the challenged claims do not require the distance determination to be based on voice analysis; they require the user-to-device distance to be calculated by a proximity sensor. PO admits Shin discloses this. (*Id.*) Second, even if the claims were to require

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.)

There can be no legitimate dispute that, under any construction, Shin discloses analyzing the user’s voice to classify it based on distance.

C. PO’s Remaining Arguments Fail.

First, PO argues that Shin does not disclose generating different output sentences (POR, 32), but this argument is based entirely on PO’s voice classification

argument and fails for the same reasons. (*Supra* §II.A.)

Second, PO questions whether Shin’s Tables 1 and 3 “are compatible with one another[.]” (POR, 33.) Shin expressly discloses that the embodiments described in Tables 1 and 3 may be used together. (*E.g.*, EX-1003 ¶¶[0049] (disclosing determining “the output scheme” which “may include an output volume level ... and an output amount of information”), [0053], [0140], claim 17 (disclosing “adjusting at least one of an output volume level of the content ... or an output amount of information of the content based on the distance between the user and the electronic device”); *see also id.* ¶¶[0183]-[0184]; Pet., 23; EX-1202 ¶23.) Petitioners’ expert testified that a POSITA would have considered Shin’s disclosure “regarding the output information and output volume together” (EX-1002 ¶88), and PO’s expert did not dispute that.

Finally, PO complains that “[t]he Petition focuses on functions without identifying what *component* allegedly accomplishes those functions.” (POR, 33-34.) Not so. The Petition consistently identifies the components accomplishing the claimed functions, which are the same components disclosed in the ’337 patent. (Pet., 9 (proximity sensor), 11-18 (processor/CPU).)

D. Shin Renders the Claims Obvious.

Shin discloses each limitation of the challenged claims. (*Supra* §§II.A-II.C; Pet., 8-30.) Because “anticipation is the epitome of obviousness,” Shin also renders

the claims obvious. *Jones v. Hardy*, 727 F.2d 1524, 1529 (Fed. Cir. 1984).

PO admits it has no obviousness-specific arguments beyond those discussed above. (POR, 35 (“[n]o party asserts” additional differences between Shin and the claims).) Instead, PO complains that the Petition “improperly conflates anticipation and obviousness.” (*Id.*) Not so. The Petition recognizes that, in an obviousness analysis, the question “is not merely what the references expressly teach but what they would have suggested” to a POSITA. *Merck & Co. v. Biocraft Labs., Inc.*, 874 F.2d 804, 807-08 (Fed. Cir. 1989). Thus, to the extent Shin fails to disclose (and thus anticipate) any claim, Shin certainly suggests (and thus renders obvious) that claim for substantially the same reasons. (*See* Pet., 30.)

PO’s additional arguments relate only to a modification that no party argues is necessary. (POR, 35-37.) As the Board recognized, where “no modifications are required,” “no showing of a motivation to modify ... or a reasonable expectation of success is required.”⁵ (DI, 18.) Thus, PO’s arguments are not relevant.

Importantly, all of Petitioners’ obviousness arguments are supported by Dr. Stern’s testimony, which is not disputed by PO’s expert. (*See* EX-2013 ¶84.)

III. GROUND 3: SHIMOMURA

PO’s arguments regarding Shimomura largely mirror those regarding Shin

⁵ Contrary to PO’s assertion, the Board did not find that “the Petition provides no basis for modifying Shin,” only that one was not required. (*Compare* POR, 37 with DI, 18.)

and should be similarly rejected.

A. Shimomura Discloses Voice Classification.

PO argues that “Shimomura does not disclose any manner of classifying voices” because it does not use the word “voice.” (POR, 38-39.) But Shimomura expressly discloses analyzing and classifying “voice” and “speech.” (Pet., 38 (citing Fig. 4); EX-1004, Fig. 4 (“voice recognition”).) And the reference is not required to use the claim’s exact words. *May*, 574 F.2d at 1090; *Whitserve*, 694 F.3d at 21. There can be no dispute that Shimomura analyzes voice (e.g., speech) input (Pet., 38-39 (citing EX-1004 ¶[0040])) and classifies it based on distance (Pet., 39-40 (citing EX-1004 ¶¶[0060]-[0064], Fig. 7)). This is exactly what the claims require.

PO argues that Shimomura’s Figure 7 discloses “a calculation of distance and nothing else.” (POR, 40.) Not so. Figure 7 discloses classification of voice input based on distance. PO admits that Figure 7 discloses “classes of distances.” (POR, 39; EX-1201, 112:13-22.) What PO ignores is that it describes classes or categories of distances *from which the voice is received*. Thus, Shimomura discloses classifying a voice based on whether the user’s distance exceeds a threshold (e.g., 350cm), exactly as in the ’337 patent. (EX-1001, 8:30-34.) The top row of Figure 7 includes two discrete classifications into which the voice may fall. (EX-1004, Fig. 7; EX-1201, 112:13-113:2; EX-1202 ¶24.) And the bottom row describes the output result

of each such classification. (*Id.*)⁶

PO argues that no POSITA would have understood Shimomura’s disclosure of distance as synonymous with voice or classification. (POR, 40-41.) But Petitioners never argued that those terms are synonymous.

PO concludes that Shimomura “directly determines the output content and volume based on distance without performing the step of classifying the voice.” (POR, 41.) That is incorrect. Shimomura uses distance to classify the received voice as coming from less than or more than a threshold distance. The result is used to determine the content and volume of the output. Again, this is precisely what the ’337 patent describes and claims. (*Id.*; EX-1001, 8:30-34; EX-1202 ¶25.)

B. Shimomura Discloses Voice Analysis.

PO argues that Shimomura’s distance calculation does not “involve[] any analysis of the user’s voice” because Shimomura uses a camera to detect distance.⁷ (POR, 43-44.) That argument is irrelevant. Nothing in the claims requires distance to be determined based on a voice analysis. Indeed, the ’337 patent never describes determining distance based on any analysis of the voice signal, only with a proximity

⁶ The “multiple users” hypothetical PO discusses for Shin, and argues is equally applicable for Shimomura (POR, 40), is irrelevant for the reasons above. (*Supra* §II.A.)

⁷ PO argues that “Amazon’s expert admitted to the same” but then cites testimony that does not discuss Shimomura or determining user-to-device distance at all. (POR, 43 (citing EX-2015, 45:3-17, 53:17-22, 54:1-6 (all discussing whisper detection)).)

sensor. (EX-1001, 8:20-34.) Thus, Shimomura describes calculating the user-to-device distance in the same way as the '337 patent, and also describes classifying voice based on that distance in the same way, i.e., whether the distance exceeds a threshold. (*Id.*; *id.*, claim 1; EX-1004 ¶[0093]; Pet., 37-40; EX-1202 ¶26.) Thus, PO's attempt to manufacture some difference between Shimomura and the '337 patent fails.

C. Shimomura Renders the Claims Obvious.

1. Limitation 1[h]

PO admits that Shimomura discloses varying output volume based on the user's distance. (POR, 45-46.) Petitioners argued that a POSITA would have been motivated to adjust Shimomura's volume change table, Figure 9, to extend the ranges so they overlapped with Shimomura's output change table, Figure 7. (Pet., 47-53.) PO makes two incorrect arguments in opposition.

First, PO argues that “the Petition never establishes that Shimomura disclosed use of the tables in conjunction with one another.” (POR, 45-46.) But Shimomura's device receives a “character string” of potential output, and based on the distance, classifies the input, determines whether to alter the output string, and then passes the resulting (potentially altered) string to be output. (EX-1004 ¶¶[0059]-[0066], Figs. 7-8; Pet., 43-46.) Then, it again uses the distance-based classification to determine whether to alter the output volume. (EX-1004 ¶¶[0066]-[0070], Figs. 9, 13; Pet., 46-

47.) Shimomura explicitly discloses that the first process is the input to the second; thus, they are used in conjunction. (EX-1202 ¶27.) PO’s caselaw deals with situations where embodiments were combined, unlike here. (POR, 45.)

Second, PO argues that Shimomura teaches away from Petitioners’ modifications. (POR, 47-51.) PO argues that a POSITA would not have used the same volume and output thresholds because there exists some distance at which “further increase in volume is not possible or not effective” and at close distances, “transforming the speech character string ... would have been unnecessary, unnatural, and potentially annoying[.]” (*Id.*, 48-50.) According to PO, “a POSITA would have understood Shimomura’s choice to set different distances as discouraging a POSITA from changing them to be the same.” (*Id.*, 50.) There are several flaws with this argument.

First, it is legally erroneous. Merely because a reference teaches accomplishing something one way does not discourage a POSITA from other ways. *DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 567 F.3d 1314, 1327 (Fed. Cir. 2009) (“express[ing] a general preference for an alternative invention” is not teaching away); *UCB, Inc. v. Actavis Labs. UT, Inc.*, 65 F.4th 679, 692-93 (Fed. Cir. 2023); *Global Shade Corp. v. With-U E-Commerce (Shanghai) Co., Ltd.*, IPR2023-00580, Paper 34 at 27-28 (P.T.A.B. Aug. 7, 2024). PO cites no disclosure in Shimomura that explicitly discourages Petitioners’ modification. (POR, 49-51.)

Second, PO’s argument rests on faulty, unsupported assumptions. For example, PO assumes Shimomura’s volume reaches a maximum limit 250cm from the device. (POR, 50.) And PO assumes that speech character string transformation would have been undesired at distances closer than 350cm. (*Id.*, 49 (citing EX-2013 ¶104, citing nothing in support).) Shimomura, in contrast, discloses only that there is some limit to the output volume (EX-1004 ¶[0072]) and that transforming the character string allows “the division of meaning [to] be clearly understood” (*id.* ¶[0057]). Nothing in Shimomura suggests that Petitioners’ modifications would be impossible or undesired.

Third, PO’s argument fails to address the motivations Petitioners provided. (Pet., 48-50; EX-1002 ¶¶164-69.)

PO also criticizes Dr. Stern, arguing that he could not provide any motivations without knowing “the purpose for the initial numbers.” (POR, 50-51.) But Dr. Stern’s motivations apply regardless of the rationale for the original thresholds. (EX-1002 ¶¶164-69.) And, Shimomura never indicates that having different thresholds is important. (*Id.*)

2. Limitations 1[f]-1[g]

PO argues that Shimomura does not disclose generating the claimed output sentences for two reasons. First, PO argues that the omission of “ne” is a change in intonation and “not an omission of information.” (POR, 51.) Not so. Shimomura

does not specify that it is talking about “ne” when it discusses change in intonation. (EX-1004 ¶[0092].) Shimomura also discloses that “words other than ‘ne’ can be used as words to be inserted,” so Shimomura’s disclosure is not limited to what PO argues. (EX-1004 ¶[0092]; EX-1202 ¶[28].) PO’s second reason (POR, 51-52) is based on its argument regarding voice classification and thus fails. (*Supra* §III.A.)

D. The Shimomura Translation Is Accurate.

PO questions the accuracy of the Shimomura translation. (POR, 37-38.) But PO did not submit its own translation; did not identify any inaccuracies in Petitioners’ translation; did not object to its admissibility; did not depose Petitioners’ translator; and did not dispute that the translator declaration (EX-1005) attests to the translation’s accuracy. *See, e.g., Xilinx, Inc. v. Godo Kaisha IP Bridge 1*, IPR2017-00842, Paper 36 (P.T.A.B. June 7, 2018). To avoid doubt, Petitioners submit with this Reply an additional declaration from Ms. Snorteland attesting to her fluency in Japanese. (EX-1206.)

IV. GROUND 4: SHIMOMURA AND SHIN

PO argues that a POSITA would not have been motivated to combine Shin with Shimomura.⁸ (POR, 53-54.) PO admits the references could be combined (*id.*, 53) but suggests there is no reason to do so, ignoring the many undisputed reasons

⁸ PO refers to earlier arguments (POR, 52), which are unavailing for the reasons above. (*Supra* §§II-III.)

Petitioners provided. (Pet., 59-64.) For example, a POSITA would have been motivated to rely on Shin’s classifications, which occur at the same distance ranges, to modify Shimomura’s classifications to likewise occur at the same distance ranges. (Pet., 60-61.) Similarly, a POSITA would have relied on Shimomura’s disclosure that its invention is not limited to the disclosed embodiments. (*Id.*, 61.) Neither PO nor its expert dispute these motivations.

PO relies on unsupported expert testimony to suggest that the combination would be “undesirable.” (POR, 53.) But as discussed, Shimomura does not teach away from Petitioners’ proposed adjustments. (*Supra* §III.C.1.) Further, PO erroneously suggests that “[n]either Shin nor Shimomura discloses classifying a voice into one of two categories.” (POR, 53.) But at least Shimomura’s Figure 7 clearly does so:

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

(EX-1004, Fig. 7; EX-1201, 112:13-113:2.) Thus, the Board should reject PO’s unsupported criticisms of Ground 4.

V. PO’S ADDITIONAL ARGUMENTS ARE UNAVAILING.

A. A POSITA Would Have Reasonably Expected Success.

PO argues that Shin and Shimomura describe simpler systems than the ’337 patent and thus a POSITA would not have been motivated to modify them. (POR, 55.) But the references disclose classifying voice input by distance in the same way, and their simplicity supports their combination. (EX-1202 ¶29.)

PO also argues that the Petition never addresses why a POSITA would have combined or modified the references and that its reasonable-expectation-of-success arguments are “[p]erfunctory.” (POR, 59-60.) But the Petition provided many motivations to combine and modify references and explained why a POSITA would have reasonably expected success. (*E.g.*, Pet., 48-50 (providing six motivations to combine Shimomura and Shin and explaining the “trivial modification” required).)

Further, PO argues that the Petition’s motivations to combine are “generic” and “lack any particular meaning.” (POR, 60-62.) PO does so by stripping the reference-specific portions of the arguments away and then arguing that they are not reference-specific. (*See id.*, 61-62; *compare, e.g., id.*, 61 with Pet., 49.) The Board should reject this circular argument as it routinely does others based on the same inapposite caselaw PO cites.⁹ *E.g.*, *ResMed Corp. v. Cleveland Med. Devices, Inc.*, IPR2025-00159, Paper 10 at 37 (P.T.A.B. June 12, 2025); *Luxottica of Am. Inc. v.*

⁹ And as the Board already has in this case. (*See* POPR, 36-38.)

E-Vision Optics, LLC, IPR2024-01069, Paper 33 at 54-55 (P.T.A.B. Dec. 16, 2025).

B. The Petition Applies the *Graham* Factors.

PO again alleges that the Petition fails to properly apply the *Graham* factors. (POR, 56-58.) This argument has already been rejected twice. (Paper 6 (PO Discretionary Denial Request), 37-42; POPR, 31-34.) It should be rejected again for the same reasons. (*E.g.*, Paper 7, 27-30.)

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).)

C. PO’s Remaining Arguments Fail.

PO repeats arguments that have already been rejected. (POR, 63-65; *see* Paper 6, 32-37; POPR, 38-43.) The Board should reject them again for the same reasons. (*See, e.g.*, Paper 7, 21-27.) In particular, PO’s argument that “[t]he Petition and expert declaration each repeatedly state contradictory alternative positions” (POR, 63) is incorrect because the statements PO identifies are not contradictory—

a teaching may be both known and obvious. Furthermore, PO's argument that Dr. Stern's declaration should be "given little or no weight" (POR, 64) is unavailing because the caselaw provides that a declaration, like Dr. Stern's, that discloses underlying facts is entitled to full weight.¹⁰ *E.g., Hum Indus. Tech., Inc. v. Amsted Rail Co., Inc.*, IPR2023-00540, Paper 10 at 37 (P.T.A.B. Aug. 11, 2023).

VI. CONCLUSION

Petitioners respectfully request the Board cancel the challenged claims.

Respectfully submitted,

MORGAN, LEWIS & BOCKIUS LLP

March 3, 2026

/ Colin B. Heideman /

Colin B. Heideman (Reg. No. 61,513)

Jeremy A. Anapol (Reg. No. 75,686)

Christie R.W. Matthaei (Reg. No. 62,933)

Nathan D. Reeves (Reg. No. 77,806)

Logan P. Young (Reg. No. 79,294)

MORGAN, LEWIS & BOCKIUS LLP

Counsel for Petitioners

AMAZON.COM, INC.,

AMAZON.COM SERVICES LLC, and

AMAZON WEB SERVICES, INC.

¹⁰ Moreover, to the extent Dr. Stern's declaration "mirrors" the Petition (POR, 64), Mr. McAlexander's declaration mirrors PO's Response. (*Compare* POR with EX-2013.)

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that this **PETITIONERS' REPLY TO PATENT OWNER'S RESPONSE** contains 5,558 words according to the word-processing program used to prepare this paper. The foregoing word count complies with the 5,600 word type-volume limit specified by 37 C.F.R. § 42.24(c)(1).

March 3, 2026

/Colin B. Heideman /
Colin B. Heideman (Reg. No. 61,513)
MORGAN, LEWIS & BOCKIUS LLP

CERTIFICATE OF SERVICE

I hereby certify that a true and correct copy of the foregoing **PETITIONERS’
REPLY TO PATENT OWNER’S RESPONSE** is being served on March 3, 2026,
via e-mail on counsel of record for Patent Owner at the address below:

Steven J. Reynolds
Kevin H. Sprenger
Chandran B. Iyer
Erin Hadi
Louay Meroueh

Daignault Iyer LLP
8229 Boone Blvd., Suite 450
Vienna, VA 22182
sreynolds@daignaultiyer.com
ksprenger@daignaultiyer.com
cbiyer@daignaultiyer.com
ehadi@daignaultiyer.com
lmeroueh@daignaultiyer.com

DI-SoundClear@daignaultiyer.com

March 3, 2026

/ Colin B. Heideman /
Colin B. Heideman (Reg. No. 61,513)
MORGAN, LEWIS & BOCKIUS LLP