

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

**PATENT OWNER'S RESPONSE TO PETITION FOR
INTER PARTES REVIEW OF UNITED STATES PATENT NO. 11,069,337**

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
II. LEVEL OF ORDINARY SKILL IN THE ART	2
III. THE ’337 PATENT	2
A. Specification	2
B. Prosecution History	3
IV. CLAIM CONSTRUCTION	5
A. Preliminary Construction of “User”	5
B. Effect of Preliminary Construction on “Classify”	6
C. Effect of Preliminary Construction on Two or More “Voices”	7
D. Effect of Preliminary Construction on “Analyze a Voice ... to Classify the Voice ... based on the Distance Between the User and the ... Device”	17
E. Petitioners Should Not Be Permitted to Take Inconsistent Claim Construction Positions	21
V. OVERVIEW OF THE ALLEGED PRIOR ART	21
A. Shin (EX-1003)	21
B. Shimomura (EX-1004)	22
C. Kristjansson (EX-1013)	22
VI. THE ’337 PATENT CLAIMS ARE VALID	23
A. Ground 1	23
1. Shin Discloses Measurement of Distance, Not Classification of Voices	23

2.	Shin Does Not Disclose Analysis of Voice for Classification	29
3.	Shin Does Not Disclose Generation of Output Based on Voice Classification.....	32
4.	Shin Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data	33
5.	Shin Does Not Disclose the Claimed Components Arranged As in the Claims	33
B.	Ground 2: Shin Does Not Render the Challenged Claims Obvious ...	35
C.	Ground 3: Shimomura Alone Does Not Render the Challenged Claims Obvious.....	37
1.	The Shimomura Translation Is Unreliable.....	37
2.	Shimomura Discloses Measurement of Distance, Not Classification of Voices	38
3.	Shimomura Does Not Disclose Analysis of Voice for Classification.....	42
4.	Shimomura Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data	44
5.	Shimomura Does Not Disclose Generation of Output Based on Voice Classification	51
D.	Ground 4: Shimomura in View of Shin Does Not Render the Challenged Claims Obvious.....	52
1.	The Combination of Shimomura and Shin Discloses Measurement of Distance, Not Classification of Voices.....	52
2.	The Combination of Shimomura and Shin Does Not Disclose Analysis of Voice for Classification	52
3.	The Combination of Shimomura and Shin Does Not Disclose Generation of Output Based on Voice Classification.....	52

4.	The Combination of Shimomura and Shin Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data ...	52
E.	Ground 5: The Combination of Shin, Shimomura, and Kristjansson Does Not Render the Challenged Claims Obvious	54
1.	The Combination of Shin, Shimomura, and Kristjansson Discloses Measurement of Distance, Not Classification of Voices.....	54
2.	The Combination of Shin, Shimomura, and Kristjansson Does Not Disclose Analysis of Voice for Classification	54
3.	The Combination of Shin, Shimomura, and Kristjansson Does Not Disclose Generation of Output Based on Voice Classification.....	54
4.	The Combination of Shin, Shimomura, and Kristjansson Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data.....	55
F.	Grounds 2-5: The Petition's Obviousness Analysis is Legally Deficient	55
1.	Illogical Modifications.....	55
2.	Failure to <i>Apply</i> Second Graham Factor.....	56
3.	No Reason for the Combinations.....	59
4.	Obviousness Arguments Are Perfunctory	59
5.	The Federal Circuit Has Rejected the Proffered Motivations to Combine	60
6.	Improper Circumvention of Obviousness and Anticipation.....	63
7.	Failure to Provide Fair Notice.....	63
8.	Expert Declaration Inseparable from Attorney-Argument	64
9.	Modifying Outputs Does Not Disclose Classification of the Claimed Inputs	64

VII. CONCLUSION.....	65
----------------------	----

TABLE OF AUTHORITIES**Page(s)****Cases**

<i>A/S v. Coolit Sys. Inc.</i> , No. 19-cv-00410-EMC, 2022 U.S. Dist. LEXIS 246288 (N.D. Cal. Sep. 11, 2022)	60
<i>Abiomed, Inc. v. Maquet Cardiovascular, LLC</i> , IPR2017-01204, Paper 9 (PTAB Oct. 23, 2017)	33, 34, 45
<i>ActiveVideo Networks, Inc. v. Verizon Communs., Inc.</i> , 694 F.3d 1312 (Fed. Cir. 2012)	62
<i>Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH</i> , IPR2019-01469, Paper 6 (PTAB Feb. 13, 2020)	4
<i>In re Anova Hearing Labs, Inc.</i> , 809 F. App'x 840 (Fed. Cir. 2020)	2, 53
<i>In re Aoyama</i> , 656 F.3d 1293 (Fed. Cir. 2011)	11
<i>Apple Inc. v. Samsung Elecs. Co., Ltd.</i> , 839 F.3d 1034 (Fed. Cir. 2016)	57
<i>Arctic Cat Inc. v. Bombardier Rec. Prods.</i> , 876 F.3d 1350 (Fed. Cir. 2017)	56
<i>Broadcom Corp. v. Emulex Corp.</i> , 732 F.3d 1325 (Fed. Cir. 2013)	60
<i>Dystar Textilfarben GmbH v. C.H. Patrick Co.</i> , 464 F.3d 1356 (Fed. Cir. 2006)	56
<i>Facebook, Inc. v. Windy City Innovations, LLC</i> , 973 F.3d 1321 (Fed. Cir. 2020)	64
<i>Graham v. John Deere Co.</i> , 383 U.S. 1 (1966)	<i>passim</i>

<i>Hayashi v. Ozawa</i> , No. 17-cv-2558 (AJN), 2018 U.S. Dist. LEXIS 53559 (S.D.N.Y. Mar. 19, 2018).....	38
<i>Hulu, LLC v. Sound View Innovations, LLC</i> , IPR2018-00582, Paper 34 (PTAB Aug. 5, 2019).....	62
<i>KSR Int’l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007).....	36, 57
<i>Kurotaki v. United States</i> , No. 22-00063	38
<i>Markman v. Westview Instruments</i> , 517 U.S. 370 (1996).....	11
<i>Metalcraft of Mayville, Inc. v. Toro Co.</i> , 848 F.3d 1358 (Fed. Cir. 2017)	56
<i>Net MoneyIN, Inc. v. VeriSign, Inc.</i> , 545 F.3d 1359 (Fed. Cir. 2008)	33, 45
<i>Palo Alto Networks, Inc. v. Centripetal Networks, LLC</i> , 122 F.4th 1378 (Fed. Cir. 2024)	2, 53
<i>Pfizer Inc. v. Sanofi Pasteur Inc.</i> , 94 F.4th 1341 (Fed. Cir. 2024)	57
<i>ProMOS Techs., Inc. v. Samsung Elecs. Co.</i> , 809 F. App’x 825 (Fed. Cir. 2020)	<i>passim</i>
<i>Thorner v. Sony Comput. Entm’t Am. LLC</i> , 669 F.3d 1362 (Fed. Cir. 2012)	10
<i>WBIP, LLC v. Kohler Co.</i> , 829 F.3d 1317 (Fed. Cir. 2016)	56
<i>Xerox Corp. v. Bytemark, Inc.</i> , IPR2022-00624, Paper 9 (PTAB Aug. 24, 2022).....	64
<i>Yita LLC v. MacNeil IP LLC</i> , 69 F.4th 1356 (Fed. Cir. 2023)	31

Other Authorities

37 C.F.R. § 42.23(b)	31
37 C.F.R. § 42.120	1
MPEP § 2143	57, 63

PATENT OWNER'S EXHIBIT LIST

Exhibit No.	Description
2001	Docket Printout of CA No. 1:24-01283 (EDVA)
2002	Order from Judge Anthony Trenga, dated January 10, 2025 (Dkt. 84)
2003	Scheduling order, dated September 23, 2024 (Dkt. 34)
2004	Final Scheduling Order, dated October 18, 2024(Dkt. 55)
2005	Judge Anthony Trenga's median time to trial statistics
2006	Excerpts from the Motion to Stay hearing transcript, dated March 26, 2025
2007	Joint proposed scheduling order (Dkt. 48)
2008	Reply in Support of Motion to Stay, dated March 20, 2025 (Dkt. 129)
2009	Full prosecution history of U.S. Patent No. 11,069,337
2010	TransPerfect Blog – The Evolution of AI in Language Processing and Its Impact on IP/Patents
2011	Median Time to Disposition in Cases Terminated After Hearing or Submission, UNITED STATES COURT OF APPEALS FOR THE FEDERAL CIRCUIT, available at https://www.cafc.uscourts.gov/wp-content/uploads/reports-stats/FY2024/MedianDispositionTime-Merits.pdf .
2012	Declaration of Anderson To
2013	Declaration of Patent Owner Expert Joseph C. McAlexander
2014	Curriculum Vitae of Joseph C. McAlexander
2015	Transcript of Deposition of Petitioners' Expert, Dr. Richard Stern

Pursuant to 37 C.F.R. § 42.120, SoundClear Technologies LLC (“SoundClear” or “Patent Owner”) submits this Patent Owner Response to the Petition for *Inter Partes* Review of U.S. Patent No. 11,069,337 (“the ’337 patent”) filed by Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively “Amazon” or “Petitioners”). For the following reasons, challenged claims 1-5 of the ’337 patent are valid over the Petition.

I. INTRODUCTION

The Petition is based upon the flawed premise that measuring distance alone constitutes a classification of voices. The Board recognized that failure. It determined, *sua sponte*, that the claims would have to be rewritten in order for the Petition to succeed. But the claim construction proposed by the Board is legally erroneous. It directly contradicts the plain meaning of related terms in the claim, including “voice” and “classification.”

That said, even under the Board’s preliminary construction, none of the prior art discloses the full context of the claims. None demonstrates a classification of voices, as opposed to merely measuring distance or volume. And none does so based upon an analysis of a voice.

The Petition also fails as a matter of law. It fails to apply the *Graham* factors in a full reasoned analysis, with rational underpinnings, as mandated by Supreme Court law. Even beyond that, the motivations to combine are incomplete, curt, and

generalized. They fall far below the standard: “The factual inquiry whether to combine references must be thorough and searching, and the need for specificity pervades [the Federal Circuit’s] authority on the Board’s findings on motivation to combine.” *Palo Alto Networks, Inc. v. Centripetal Networks, LLC*, 122 F.4th 1378, 1384 (Fed. Cir. 2024) (vacating when “the Board erred by not addressing the evidence and the argument for the motivation to combine and by failing to provide an adequate explanation”); *In re Anova Hearing Labs, Inc.*, 809 F. App’x 840, 843 (Fed. Cir. 2020) (same).

II. LEVEL OF ORDINARY SKILL IN THE ART

A POSITA at the time of the invention of the ’337 patent would have: (a) a Bachelor of Science degree in electrical or computer engineering (or a related academic field), and at least two (2) additional years of work experience in the design and development of electronic devices or systems related to controlling audio output based on inputs. EX-2013, ¶32.

III. THE ’337 PATENT

A. Specification

The ’337 patent discloses a “voice-content control device” that includes “a voice classifying unit configured to analyze a voice spoken by a user.” EX-2013, ¶35 (citing EX-1001, Abstract). The focus is on user interaction with a device by speaking to it. More particularly, it relates to a device that analyzes a user’s voice.

Id. That voice includes multiple aspects, including the “content” contained in it. *Id.*

The '337 patent is not a mere measurement device. *Id.* It is not simply a “volume measurement device” or “distance determination device.” *Id.* It is focused on human voices and what makes voices different than any singular characteristic, such as volume or distance. *Id.*

Various embodiments in the '337 patent describe classifying whether a voice is a whisper or not. EX-2013, ¶37 (citations omitted). The patent describes multiple characteristics of “voice” that are used. Those include, for example, analysis of the waveform, intensity, peak frequency of the sound waves, the volume of the voice, the speaking speed of the voice, the volume ratio between the speech and wind noise, the distance between the speaker and the device, or the mel-frequency cepstrum coefficient. *Id.*, ¶¶37-38 (citations omitted).

B. Prosecution History

During prosecution, the Examiner considered Shin as “Pertinent Prior [A]rt” and allowed the claims over it. EX-2009, 176.

Shin is the lead reference in Amazon's Petition. Thus, *Becton* factor (a) favors step 1. “Previously presented art includes art made of record by the Examiner, and art provided to the Office by an applicant, such as on an Information Disclosure Statement (IDS), in the prosecution history of the

challenged patent.” *Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH*, IPR2019-01469, Paper 6, 7-8 (PTAB Feb. 13, 2020).

The Examiner listed Shin as “Pertinent Prior art” and stated that “Paragraph 37 discloses a response to the voice input is adjusted or determined according to the volume of the user input.” EX-2009, 176; EX-1003 ¶[0037].

After the Examiner issued the Final Rejection, SoundClear filed its response on March 9, 2021. EX-2009, 153. SoundClear amended Claim 1 to add the limitation of “a proximity sensor configured to calculate a distance between a user and the voice-content control device; . . .” and “based on the distance between the user and the voice-controlled device. . . .” *id.*, 154. The Examiner issued the Notice of Allowance less than two weeks later, on March 22, 2021. *id.*, 112.

Amazon's Petition misrepresents the record in this regard. It asserts that the Examiner “overlooked . . . that the device's voice response outputs *different content . . . at a different volume . . . based on the distance between the user and the device . . .*” Pet., 71 (emphasis in original). That is incorrect; the Examiner specifically asserted that “Paragraph 37 [of Shin] discloses a response to the *voice input*” that “*is adjusted or determined according to the volume of the user input.*” EX-2009, 176. Thus, the Examiner did expressly consider purported disclosures of “*different content*” at “*different volume.*” And the opening sentence of paragraph [0037] of Shin expressly mentions distance: “a user 10, who is spaced apart from

an electronic device 100 *by a specific distance.*" EX-1003, ¶[0037] (emphasis added). In fact, the entire point of that paragraph (discussing Fig. 1A) is to take distance into consideration. It describes different adjustments based upon a user being "at a standstill." It then contrasts that with Fig. 1B where the user "moves away." *Id.*, ¶[0038].

IV. CLAIM CONSTRUCTION

Neither Patent Owner nor Petitioners proposed any claim construction at the institution stage of this proceeding. Pet., 8; Patent Owner's Preliminary Response (Paper 9, "POPR"), 8. But, in the Institution Decision (Paper 12, "Dec."), the Board *sua sponte* stated that claim construction of "user" was necessary. Dec., 14.

A. Preliminary Construction of "User"

The Board's construction is unnecessary. The Board preliminarily construed the claimed "a voice spoken by a user" to encompass "one or more voices spoken by a single user, or by plural users." *Id.*, 15. The Board explained that its "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." Dec., 16 (emphasis added). The Board "invited [the parties] to fully develop this claim construction issue during trial." *Id.* That is not necessary. Patent Owner does not contest the Board's conclusion that the claims recite "one or

more users.” In certain places of its POPR, Patent Owner explained that the prior art did not distinguish between multiple users. *See, e.g.*, POPR, 16. If it cannot distinguish between two users—who will *always* have different voices—it cannot distinguish between “the type of sound or which voice [a single user] is speaking.” *Id.*, 28. Patent Owner acknowledges that the POPR could have been clearer on that front. But, with this clarification, neither party contests that the scope of the claims extend to a single user speaking in multiple voices. So, no construction of “user” is necessary.

B. Effect of Preliminary Construction on “Classify”

The preliminary construction is problematic for other reasons. It would drastically broaden the claims by vitiating important elements. The preliminary construction would convert the recited “classification” into a measurement. EX-2013, ¶¶42-43. A POSITA would have understood “classify” to require something discrete, not a mere measurement. *Id.* The claims make that clear by reciting two (or more) distinct “voices”: a first and second voice. A POSITA would not have understood “classify” to be used in the context of measuring a single, continuously changing variable (like volume or distance). *Id.* A POSITA would have understood the claims of the ’337 patent to require a classification, or distinction between, at least two separately identifiable voices. *Id.*

C. Effect of Preliminary Construction on Two or More “Voices”

The preliminary construction is also problematic in that it renders the word “voice” superfluous. EX-2013, ¶44. “Voice” does not equate to the mere incidental change in “volume” due to a measurement device being moved away from a speaker. *Id.* It is especially inapt in the context of a user who speaks at a constant volume but simply moves away from a measuring device. *Id.* Again Petitioners’ expert agrees that the concept of “voice” cannot be reduced to a simple metric like distance:

Q Does the distance between the user and the device alone tell you anything about the characteristics of the user’s voice?

MR. REEVES: Objection. Form. Outside the scope.

A In general, no.

EX-2015, 54:1-6.

Nor can the concept of “voice” be reduced to a simple metric like volume because it is tied to the concept of speech:

Q So what’s the plain and ordinary meaning of the term voice that you apply when analyzing the claims?

A The voice I consider to be the natural productions that humans make *in communicating with each other via speech* and also that humans make in talking to machines using *speech*.

Id., 28:12-18 (emphasis added);

Q I've noticed that you've been using the phrase "speech classification" and "voice classification" interchangeably. How is voice different from speech or sound?

...

A In this context I have been using those terms somewhat interchangeably. Sound encompasses a lot of nonspeech production such as audio, music, wind noise, hurricane noise. Humans can also produce, you know, nonverbal expressions such as laughter or screams, things like that. When I speak about speech and/or voice, and I do use those terms somewhat interchangeably, I'm talking about the subset of things that humans utter that *are directed toward communication of an idea or a request or something like that, normal conversation.*

Id., 31:11-32:6.

The word "voice" connotes the quality of information conveyed by a human. EX-2013, ¶45. That information is not conveyed with simple changes in volume. *Id.* It is conveyed via a multivariate notion of voice. *Id.* To be sure, the nature of a (singular) voice can change with volume. *Id.* But volume alone does not create a new voice:

Q Okay. But I do think this is within the scope of the patent. The patent talks about different voices and increasing and decreasing

volume. So I'm just asking you, does the nature of the voice change when you increase the volume of your voice?

...

A The patent -- the patent really does not address that, the nature of the changes, that do happen when you increase and decrease your voice.

Id. (citing EX-2015, 59:13-50). When asked to clarify, Petitioners' expert attempted to decline to provide an answer. *Id.* (citing EX-2015, 60:3-62:16). He stated that he would "stand" on his refusal to answer a question—even after discussing the matter with counsel—without any basis for doing so (e.g., privilege). But when pressed, despite his resistance, he did state that something *more* than volume or intensity ("other things") must change before voice changes. EX-2013, ¶45; EX-2015, 60:3-62:16. Something more is necessary before one can "discriminate between speech that's shouted from speech that's spoken quietly, *regardless of what the volume control is*" EX-2013, ¶45 (citing EX-2015, 60:3-62:16).

The preliminary construction would mean that every "chang[e in] voice volume" would be deemed to be a new voice. (e.g., a "second voice"). That interpretation would strip all meaning from the term "voice" and from the classification of multiple voices. It would transform the act of classifying spoken

voices into the measuring of changes in volume of signals received at a device.

And it would directly contradict even Petitioners' expert's testimony (quoted above).

Moreover, the preliminary construction conflates multiple terms that have plain and ordinary meanings and that are used throughout the patent. EX-2013, ¶46. A POSITA would have understood "voice," "volume," and "distance" to be distinct concepts. *Id.* In fact, the claims expressly provide for volume to change while voice remains constant. *Id.*; *see also* discussion of limitations [e] and [h] below.

The preliminary construction would also be legally erroneous, for at least four reasons.

First, a claim must be given its plain and ordinary meaning unless two exceptions apply: (1) lexicography or (2) disavowal. *Thorner v. Sony Comput. Entm't Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012). Neither exception applies here. Neither Petitioners nor the Board assert that one does.

Second, terms should not be rendered superfluous. *ProMOS Techs., Inc. v. Samsung Elecs. Co.*, 809 F. App'x 825, 834 (Fed. Cir. 2020).

Third, the Federal Circuit "specifically cautioned against construing claims by first developing a broad definition of claim terms divorced from the context of the patent and then turning to the specification only to look for grounds for limiting

the claim scope.” *ProMOS*, 809 F. App’x 825 at 836-37 (citing *Phillips v. AWH Corp.*, 415 F.3d 1303, 1320 (Fed. Cir. 2005)). The court thus “rejected a proposed claim construction where it relies on an initial ‘assumption’ that a claim term has a broad ‘ordinary meaning’ unsupported by the intrinsic or extrinsic evidence.” *Id.* (citing *Ruckus Wireless, Inc. v. Innovative Wireless Sols., LLC*, 824 F.3d 999, 1003 (Fed. Cir. 2016)). The preliminary construction would have to be founded on an assumption. There is no record here to support very broad meanings of “classify” and “voice.”

That preliminary construction would be erroneous for a fourth reason as well. It is bedrock law that “the meaning and scope” of claim terms be determined *before* the claims are applied to the accused device or method or prior art.

Markman v. Westview Instruments, 517 U.S. 370, 384 (1996); *In re Aoyama*, 656 F.3d 1293, 1296 (Fed. Cir. 2011). A construction based on the prior art (e.g., in the context of Shin) would be improper—especially because it discloses a system with a different purpose and approach to varying content.

The intrinsic evidence demonstrates that voice cannot simply equate to “volume” or “distance.” Dec., 16. “When the plain meaning of a claim term does not clarify whether and how a restriction applies, [the Federal Circuit has] said that such ambiguity is to be resolved in the first instance by reviewing the claim language as a whole, the other intrinsic record, and any extrinsic evidence.”

ProMOS Techs, 809 F. App'x at 837 (citing *Uship Intellectual Props., LLC v. U.S.*, 714 F.3d 1311, 1315 (Fed. Cir. 2013)). Each of those sources confirms that the concept of a voice is distinct from any single characteristic (e.g., volume).

The claim language alone is dispositive. The claims recite “**voice**” separate from both “**volume**” and “**distance**.” For instance, claim 1 follows:

1. A **voice**-content control device, comprising:

- [a] a proximity sensor configured to calculate a **distance** between a user and the **voice**-content control device;
- [b] a **voice** classifying unit configured to analyze a **voice** spoken by a user and acquired by a **voice** acquiring unit to classify the **voice** as either one of a first **voice** or a second **voice** based on the **distance** between the user and the **voice**-content control device;
- [c] a process executing unit configured to analyze the **voice** acquired by the **voice** acquiring unit to execute processing required by the user;
- [d] a **voice**-content generating unit configured to generate, based on content of the processing executed by the process executing unit, output sentence that is text data for a **voice** to be output to the user; and
- [e] an output controller configured to adjust a sound **volume** of **voice** data obtained by converting the output sentence thereinto, wherein

- [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**,
- [h] wherein 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, claim 1 (annotated). The distinction between voice and volume is unmistakable. EX-2013, ¶47. The claim recites measuring a distance (limitation [a]) and then three limitations ([b]-[d]) regarding classifying voice, analyzing voice, and generating content for a voice. *Id.* All are conducted without reciting “volume.” *Id.* The claim recites that volume is a characteristic that can be “adjusted.” *Id.* Volume is not “classified.” *Id.* And that volume adjustment is not recited until a separate limitation is introduced (limitation [e]). *Id.* There is thus no basis to conflate concepts that the claim recites separately. *Id.*

Further, the Board's preliminary construction disregards the plain language of limitations [e] and [h]. Those limitations recite adjusting volume of *one* output voice. Volume changes continuously but *the* output voice remains constant.

The full context of the claim confirms that interpretation. *Id.*, ¶48. Voice is “classified.” *Id.* It is a discrete concept. *Id.* The incoming voice can either be one voice or one or more other voices. *Id.* There is no middle ground. *Id.* It is not a single variable that can have any value (on a continuous scale). *Id.* It is not something that is understood to be calculated, measured, or “adjusted” like volume. *Id.* Further, the claim mandates that the classification include an assessment of distance. *Id.* Whether or not the classification is based on volume is of no moment. *Id.* There is also no indication that the patentee recited distance as a mere proxy for volume, especially when volume could be measured directly. *Id.* Petitioners’ expert admits that while distance may sometimes be sufficient to classify voice, it is not always sufficient. *Id.* (citing EX-2002, 48:19-49:7). That is crucial. It means that distance cannot possibly be synonymous with “voice.” *Id.* It is but one aspect of “voice.” *Id.*

The specification supports that conclusion. The ’337 patent is entitled “Voice-Content Control Device, Voice-Content Control Method, and Non-Transitory Storage Medium,” and “relates to a voice-content control device, a voice-content control method, and a non-transitory storage medium.” *Id.*, ¶49 (citing EX-1001, code (54), 1:15-17). It is not, for example, merely a “volume-content control device.” *Id.* The ’337 patent discloses “a voice processing device that outputs voice.” *Id.* (citing Dec., 3 (citing EX-1001, 1:29-32, 1:39-42)). It does

not disclose a “volume processing device.” *Id.* Nor does it disclose a device that outputs “volume.” *Id.*

The entire disclosure demonstrates what any POSITA would have understood: that voices contain many *characteristics*. *Id.*, ¶50. Those characteristics include, for example, volume, frequency, intensity, slopes of peak frequency, content, speaking speed, volume ratio with wind noise, or others. *Id.* (citing EX-1001, 7:64-8:34). The specification expressly defines voice as incorporating each of those characteristics, not just volume. *Id.* It is a discrete concept, whereas characteristics like volume, frequency, intensity, and slopes of frequencies are continuously varying attributes. *Id.* Those attributes can, for example, *decrease* or can be *too much*.

In that case, the output voice itself may be *decreased* in sound volume, but if the output voice is *decreased* in sound volume *too much*, the output voice may be hard to be heard by the user himself and the user may be unable to understand the content of the voice. *Id.* (citing EX-1001, 1:34-38 (emphasis added)). The above quote again confirms that a single voice (*the* output voice) can have multiple volumes (can be increased or decreased to any value). *Id.*

The specification even gives a specific example. *Id.*, ¶51 It describes classifying “the voice V1 to the second voice V1B *when the voice V1 is determined to be a whisper as a result of analysis,*” and “to the first voice V1A

when the voice V1 is determined not to be a whisper.” Id. (citing EX-1001, 7:51-63 (emphasis added)). It also describes numerous other characteristics of “voice,” such as “feature values,” and how they can be used to classify voices. *Id.*, ¶¶51-53. And Petitioners’ expert admits that the specification’s description of the above “feature values” are for purposes of determining whether a user’s voice is a whisper or not a whisper. EX-2015, 36:7-37:6. Notably, each can be used alternatively or in different combinations. *Id.* But no such changes alter the nature of what is being classified: voice. EX-2013, ¶53. Certainly, a POSITA would not have understood use of a measurement to mean that any single characteristic defines the to-be-classified voice. *Id.* 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. *Id.*

This point is crucial. Just because a POSITA might choose to use one variable to classify voice does not mean that the classification will be perfectly effective. *Id.*, ¶54. Nor does it mean that any error associated with such single-variable classification would have transformed one speaking voice into a new voice. *Id.* Otherwise, all claims, in any patent, would be broadened by the extent to which an invention might operate with some error in the real world. The claims of the ’337 patent mandate that a classification be made and that the classification

render a determination as to at least two discrete voices. *Id.* Simply detecting volume is not enough. *Id.*

The Board's reasoning recognizes as much. *Id.*, ¶55. It acknowledges that at minimum a voice is something a user "perform[s]." *Id.* (citing Dec., 16). A user that speaks at the same volume while walking away from a device has not changed any aspect of their performance. The user has not even changed the volume of their own voice. In that instance, the device does not detect any aspect of user performance. It is the volume that the device receives that is different. In that way, the preliminary construction would defeat the entire purpose of the patent. The classification unit would simply be a volume measurement.

D. Effect of Preliminary Construction on "Analyze a Voice ... to Classify the Voice ... based on the Distance Between the User and the ... Device"

Limitations 1[b], 4[c], and 5[c] expressly require (1) a classification that must be based on (2) an analysis of the acquired voice and (3) a distance, which is calculated (*e.g.*, by a proximity sensor) in limitations 1[a], 4[a], and 5[a]. *Id.*, ¶56.

With respect to requirement (3), the antecedent basis for "the distance between the user and the voice-content control device" is found in limitations 1[a], 4[a], and 5[a], which respectively recite:

- Limitation 1[a]: "a proximity sensor configured to calculate a distance between a user and the voice-content control device"

- Limitations 4[a] and 5[a]: “*calculating a distance between a user and a voice-content control device; acquiring a voice spoken by a user*”

EX-1001. The claims thus expressly provide that the purpose of the analysis of the acquired voice (requirements **(1)** and **(2)**) is to “classify the voice” as either one of a first voice and a second voice, and not for calculating “the distance between the user and the voice-content control device” (**requirement (3)**). The latter, as explained above, is calculated (*e.g.*, by a proximity sensor) in limitations 1[a], 4[a], and 5[a].

The specification consistently describes analyzing a user's voice for purposes of determining whether the user's voice is a whisper and classifying the user's voice on that basis. EX-2013, ¶57 (citing EX-1001, 7:51-8:34). The specification describes calculating distance with a proximity sensor. *Id.* (citing 8:20-26). The specification further provides that “[t]he voice classifying unit 38 ... can perform the classification *by combining* the classification methods explained” therein, including by combining voice analyses (*e.g.*, to determine whether the user's voice is a whisper) with distance calculation (*e.g.*, by a proximity sensor), as required by the plain language of the claims. *Id.* (citing 8:56-59, 7:64-8:34).

Further, the specification discloses the “distance between the user H and the voice-content control device 1” as a potential “feature value” for determining whether the user's voice is a whisper and classifying the user's voice based on the

same. EX-2013, ¶58. A POSITA would have recognized that such distance (*i.e.*, between the user and the device) alone is not a good indicator of whether the user's voice is a whisper. EX-2013, ¶58; EX-2015, 36:7-37:6. In view of the specification's disclosure, a POSITA further would have understood that distance between the user and the device can be used in conjunction with other "feature value[s]" (e.g., peak frequency equal to or higher than a predetermined intensity, slope of the peak in the frequency spectrum, voice volume, speaking speed, volume ratio between speech and noise, mel-frequency cepstrum coefficient) to determine whether the user's voice is a whisper for purposes of classifying the user's voice. *Id.* (citing EX-1001, 7:64-8:34).

Thus, in view of (and consistent with) the specification, a POSITA would have understood the plain meaning of limitations 1[b], 4[c], and 5[c] to require the claimed (1) voice classification be based on (2) analysis of the acquired voice and (3) *distance* calculated (*e.g.*, by a proximity sensor) in limitations 1[a], 4[a], and 5[a]. *Id.*, ¶59. The Board's preliminary construction would "*only* require[] ... a single user who may perform both a first and second voice *depending [solely] upon* the single user's changing *distance* ... from the voice-content control device." Dec., 16. It is thus erroneous for at least two reasons.

First, the Board's preliminary construction would only require classifying the acquired voice as either a first or second voice based on distance between a

user and the device. That is, irrespective of whether *any* analysis of the acquired voice is performed. Such a construction renders meaningless the concept of a **(1)** classification (turning it into measurement of a continuous variable) and of **(2)** limitations 1[b], 4[c], and 5[c] (that the acquired voice be analyzed for purposes of voice classification). It reads out requirement **(3)** and rewrites limitations 1[b], 4[c], and 5[c] as follow. In short, even setting aside the problem of reading out the **(1)** classification, the preliminary construction would *still* drastically modify the claim:

- **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;

EX-1001, claims 1, 4-5.

Second, to the extent the Board's preliminary construction would conflate the two separate requirements within limitations 1[b], 4[c], and 5[c]. *See* Dec., 16 (“[O]ur 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.” (emphasis added)). As explained above, the expressed purpose of the claimed “analyze the acquired voice [spoken by a user]” in requirement (2) is “to classify the acquired voice” (in requirement (1)) and not to determine “*the distance*” in requirement (3).

E. Petitioners Should Not Be Permitted to Take Inconsistent Claim Construction Positions

Petitioners should not be permitted to take conflicting claim construction positions for purposes of invalidity (in this IPR) and infringement (in district court). In the parallel district court case, Petitioners stated that it is a matter of “when,” not “if,” it will seek to construe the claims. EX-2006, 3-4, 7. Petitioners will seek to invalidate the patents by applying a broad scope of the claims before this Board. They will then turn around and seek to avoid infringement by seeking to have the district court issue narrowing constructions.

V. OVERVIEW OF THE ALLEGED PRIOR ART

A. Shin (EX-1003)

Shin discloses an electronic device for providing content in response to user input. EX-2013, ¶¶60-61 (citing EX-1003). It discloses adjusting content based on “least one of a speech rate of *the* speech, a volume of *the* speech, or a keyword included in *the* speech . . .” *Id.* (citing EX-1003, ¶[0010] (emphasis added)). The Petition agrees: Shin operates “*based on* the user-to-device distance.” *Id.* (citing Pet., 19 (emphasis added)).

The Patent Office issued the '337 patent over Shin. The Examiner considered Shin as “Pertinent Prior Art.” EX-2009, 176. The Examiner cited a portion of Shin that is redundant of all of the other disclosures on which Petitioners rely—paragraph [0037]. The Examiner specifically allowed the challenged claims over Shin.

B. Shimomura (EX-1004)

Shimomura is concerned only with distance and the volume output based on distance. EX-2013, ¶¶62-64. It thus lacks all voice classification limitations, especially those based on an analysis of the voice and distance:

[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

Id. (citing EX-1004, 1 (emphasis added), ¶[0012]). Shimomura also discloses a “change[] in intonation.” *Id.* (citing EX-1004, ¶[0092]).

C. Kristjansson (EX-1013)

Kristjansson discloses a device that uses distance to determine sound pressure where the user is located. *Id.*, ¶¶65-66 (citing EX-1013, Abstract).

VI. THE '337 PATENT CLAIMS ARE VALID

Each ground is addressed in connection with the specific limitations at issue. But, in large part, the failures of the prior art are simple. The Examiner considered Amazon's lead reference (Shin) and issued the patent over it. Shimomura does not add anything the Examiner did not already consider. None of the prior art discloses a classification of voices, as opposed to measuring distance or volume. And none do so based upon an analysis of a voice. That is true under the preliminary construction and plain and ordinary meaning of "user."

A. Ground 1

1. Shin Discloses Measurement of Distance, Not Classification of Voices

The Petition relies upon Shin's measurement of distance to meet all limitations of the claims. But Shin fails to do so. As discussed in Section IV.B-C above, limitations 1[b], 4[c], and 5[c] require classes of *voices* (i.e., "a first *voice* and a second *voice*"). None of the asserted grounds and prior art therein discloses classifying a voice spoken by a user "as either a first voice [and/or] a second voice," as claimed and described in the specification.

The Petition acknowledges that Shin "adjusts the amount of information provided to the user *based on* the user-to-device distance." Pet., 19. The Petition does not purport to identify any express disclosure of basing *any* function or output on "voice." It simply removes adjustment of information, and adds claim language,

in conclusory fashion (the portion in italics): “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.* ¶68 (citing Pet., 13 (1st full ¶) (emphasis added) (citing only expert declaration)).

The Petition also contradicts Shin. The only express disclosure of “voice” the Petition identifies in Shin that relates to processing voice is in connection with “generat[ing] . . . content.” *Id.* (citing Pet., 13 (1st full ¶, 2d sentence)). That is, garden variety natural language processing to determine user intent. *Id.* None of the sentences that follow in that section (VII.A.3) point to any express disclosure of “voice” or “classification” in Shin. The only exception occurs in Shin ¶[0051]. But it only mentions “classification” of “content.” *Id.* (citing EX-1003, ¶[0051]). And Shin mentions that classification of content solely in the context of output, and not voice input. *Id.* (citing EX-1003, Table 1, ¶[0051]). All mentions of “voice” beyond the above-cited sentence in the Petition (including those with quotes around them) are taken from Petitioners’ expert or the ’337 patent. Neither the Petition nor Petitioners’ expert provides any reasoned analysis supporting the conclusion that Shin classifies voices.

The Petition then turns to Table 3 at the bottom of page 13. EX-2013, ¶69. It includes an annotated version of the table, where its expert added the “First Voice” and “Second Voice” labels. *Id.* (citing EX-1002, ¶64). That is the sole basis for the

Petition. But annotations aside, Table 3 discloses only a calculation of distance and nothing else:

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)

Id. (citing EX-1003, Table 3).

The determination of distance in Table 3 cannot expressly disclose a classification because it is a measurement of a continuous variable, not discrete classification. *Id.*, ¶70. Nor can it expressly disclose an analysis of a “voice,” which is not at all synonymous with “distance.” *Id.* One example is sufficient to demonstrate the failed reasoning here. *Id.*, ¶71. Assume two users, with two different voices, were standing the same distance from the device disclosed in Shin. *Id.* Table 3 shows that Shin would not distinguish between the voices of those users. *Id.* The only information determined by Shin would be distance. And it would be the same for each voice. *Id.*

In fact, Shin itself states that it only contemplates a single voice in connection with Table 3: “For example, the content that has the greatest output

amount of information may correspond to the most detailed original content corresponding to a voice input.” *Id.*, ¶72 (citing EX-1003, ¶[0087] (emphasis added)). And Shin confirms the distinction between “voice” and “distance.” *Id.* Shin states that “processor 120 may determine a distance between the user and the electronic device 101 based on at least one of the *volume* of the user’s speech obtained through an analysis of a *voice* input or the *distance* computed, calculated, or measured by the distance detection module 180.” *Id.* (citing EX-1003, ¶[0053]). Thus, a POSITA would never have understood Shin’s disclosure of distance to be synonymous with “voice.” *Id.* Nor would a POSITA have understood Shin’s determination of a distance to be synonymous with a classification. *Id.*

Moreover, Shin directly correlates content output to distance between the user and the device. *Id.*, ¶73. It does not classify the voice input in any way, let alone classifying the voice input as a first voice or a second voice. *Id.* Petitioners tacitly acknowledge that Shin’s classes/groupings are based on distance, not voices. Pet., 14 (“*Any of the distance ranges up to 4 meters* would satisfy the claimed ‘first voice.’ For example, a voice between 3 and 4 meters away could be considered the recited ‘first voice.’ Alternatively, *all of the distances less than 4 meters* could collectively be considered a first voice (e.g., a voice less than 4 meters away).” (citations omitted)). In contrast, as discussed in Section IV.C, the specification consistently describes the claimed first and second voice as different

classes of **voices** (e.g., a whisper voice, i.e., “an unvoiced voice spoken without vibrating the vocal chords”). EX-1001, 7:59-63.

Crucially, Shin's Table 3 does not even disclose voice classification under Petitioners' expert's own understanding. Specifically, Petitioners' expert testified that whether a classification constitutes a **voice** classification depends on the input and output of the classification. EX-2015, 23:5-24:17. However, neither the input (i.e., distance) nor output (device content) of Shin's Table 3 is the user's voice, let alone a “first voice” or a “second voice” as claimed.

Petitioners' expert also admitted that Table 3 says “nothing” other than “about the distance between the user and the electronic device”:

Q Can you looking at table 3, can you -- does it provide any information regarding the user's voice in table 3?

MR. REEVES: Objection. Form. Asked and answered.

A It has been asked and answered. I'm not sure what more I can tell you that I haven't already.

Q Well, you keep -- you keep answering in the context of classification. That's not my question, sir. My question is simply, does table 3 provide any information regarding the user's voice, not whether you can use table 3 to classify the voice?

MR. REEVES: Same objections.

A No. Just so we don't continue iterating this for the rest of the day, it tells you about the distance between the user and the electronic device and nothing else.

Id., 72:12-73:8 (emphasis altered).

When asked where classification of voices came in, he referred to the discussion in his declaration "leading up to table 3 which is referred to on page 31. . . ." *Id.*, 69:19-70:3. But that discussion (in his declaration) exactly mirrors the first full paragraph on page 13 of the Petition. And, as discussed above, that paragraph only addresses Shin's disclosure of analyzing voice to determine speech content. There is no factual basis, and no reasoned analysis, upon which to conclude that a POSITA would have understood Shin to disclose voice classification. Petitioners' expert's sworn testimony conclusively establishes that the conclusions in the Petition are circular and ultimately based solely on his say-so (*i.e.*, conclusions without reasoned analysis).

The above reasoning applies even under the Board's preliminary construction. That is, even if a new voice could theoretically be created based solely on a volume change based on distance, that would not mean that a new voice is always created simply because distance is changed. EX-2013, ¶74. The concepts of "volume" and "distance" would still be different than the concept of a "voice." *Id.* The Petition provides no basis upon which the Board could conclude

that Shin discloses the concept of determining a voice as opposed to simply making a measurement of “distance.” *Id.* And a POSITA would never have understood Shin’s disclosure of calculating a change in distance of a user, who otherwise does not change their voice at all, as any sort of voice classification—even if volume received at the device changed as well. *Id.*

2. Shin Does Not Disclose Analysis of Voice for Classification

Limitation 1[b] recites: “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.*” *Id.*, ¶75 (citing EX-1001, Claim 1). Limitations 4[c] and 5[c] are similar for purposes of this discussion. *Id.*

As discussed in Section IV, those limitations require **(1)** a classification that must be based on **(2)** an analysis of the acquired voice and **(3)** a distance, which is calculated (*e.g.*, by a proximity sensor) in limitations 1[a], 4[a], and 5[a]. *Id.*, ¶76. Shin does not disclose the claimed “analyz[e/ing] the [acquired] voice to classify the [acquired] voice.” *Id.*

The Petition asserts that Shin discloses “analyz[ing] a voice spoken by a user and acquired by a voice acquiring unit” (as recited in claim 1 and similarly recited in claims 4-5). *Id.*, ¶77 (citing Pet., 12-13). It relies on Shin’s disclosure of a processor 120 that “analyze[s] a voice input received through [the] audio input

module 151.” *Id.* (citing Pet., 12-13 (quoting EX-1003, ¶[0048]), 29-30 (alleging that Shin discloses claims 4-5 for the same reasons discussed in analysis for claim 1)). Paragraph [0048] of Shin, however, relates to *content* of speech—what a user is saying. *Id.* It does not relate at all to the distance detection disclosure, let alone any classification of voices. *Id.*

In any event, the Petition omits any discussion of what the analysis disclosed in paragraph [0048] is used for. It moves right to the output side of Shin, which is disclosed in an unrelated part of Shin’s disclosure. *Id.* It asserts that Shin’s “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.* (citing Pet., 13 (quoting EX-1003 ¶¶[0078], [0086])). The Petition presents no analysis as to how either the input or the output functions disclosed by Shin meet the claims. *Id.* It does not explain where Shin discloses analysis of the user’s voice in order “*to classify the voice* as either one of a first voice or a second voice” as claimed. *Id.*

As discussed above, the Petition relies entirely on Table 3 as disclosing the claimed classification. *Id.*, ¶78. It states that Shin discloses the claimed voice classification because Table 3 shows “providing a different amount of information based on the user-to-device distance.” *Id.* (citing Pet., 13-14). It further states that

“Shin discloses classifying the voice as a ‘first voice’ when the distance is less than 1 meter, and a ‘second voice’ when the distance is greater than 4 meters.” *Id.*

However, the Petition includes no allegation that Shin determines the user-to-device distance (as used in Table 3 to determine “[o]utput amount of content of information”) based on any voice analysis in connection with populating or using Table 3. EX-1003, Table 3. Consistent with that, under Petitioners’ mapping, Shin makes clear that voice analysis is *not* used for that purpose. *Id.*, ¶79. Shin “compute[s] a distance between a user and the electronic device” using camera and/or sensors in its “distance detection module.” *Id.* (citing Pet., 9-11). Thus, there is no express disclosure or any other reason to believe Shin discloses “analyze [the] voice ... *to classify the voice.*” *Id.*¹

Accordingly, Petitioners did not and could not demonstrate that Shin discloses “analyze a user’s voice ... to classify the voice” as recited in claim 1 (and similarly recited in claims 4-5) under both the plain meaning of the claims and

¹ Even if there were support in Shin for that position, Petitioners cannot change theory at this stage of the proceeding. 37 C.F.R. § 42.23(b); *Yita LLC v. MacNeil IP LLC*, 69 F.4th 1356, 1367 (Fed. Cir. 2023) (Board should not rely on reply arguments that the petitioner “could have presented ... its petition but chose not to” (cleaned up)).

Patent Owner's construction because no voice analysis whatsoever is involved in Shimomura's determination of user-to-device distance based by image recognition unit 41.

3. Shin Does Not Disclose Generation of Output Based on Voice Classification

Claim 1[f]-[g] recite: "generate a first output sentence as the output sentence when *the acquired voice has been classified as the first voice*, and 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*." EX-2013, ¶80 (citing EX-1001, Claim 1). Limitations 4[g] and 5[g] are similar for purposes of this discussion. *Id.* The Board's preliminary construction does not affect this limitation. *Id.*

For reasons discussed in Sections VI.A-B above, none of the asserted grounds and prior art references therein discloses or suggests classifying the acquired voice as the first and second voice as claimed. *Id.*, ¶81. For the same reasons, none of the asserted grounds discloses or suggests generating a first and second output sentence when the acquired voice has been classified as the first and second voice. *Id.*

4. Shin Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data

Claim 1 requires, *inter alia*, “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, Claim 1. Limitations 4[h] and 5[h] are similar for purposes of this discussion. The Board’s preliminary construction does not affect this limitation.

The Petition relies on two separate tables disclosed in Shin: Table 1(volume) and Table 3 (information). Pet., 21-23. The Petition does not provide any analysis supporting the assertion that the processes disclosed in connection with Tables 1 and 3 are compatible with one another or if or how a POSITA would have understood them to be used in conjunction with one another. *Net MoneyIN, Inc. v. VeriSign, Inc.*, 545 F.3d 1359, 1370-71 (Fed. Cir. 2008); *Abiomed, Inc. v. Maquet Cardiovascular, LLC*, IPR2017-01204, Paper 9, 8-12 (PTAB Oct. 23, 2017).

5. Shin Does Not Disclose the Claimed Components Arranged As in the Claims

Ground 1 is based upon anticipation, but Ground 5 demonstrates that Ground 1 is insufficient on that front. That is because the Petition does not identify where the components of claim 1 are actually disclosed in Shin. EX-2013, ¶82. The

Petition focuses on functions without identifying what *component* allegedly accomplishes those functions. *Id.* For example, the Petition describes the '337 patent's "voice classifying unit" as referring to "a CPU configured to perform the claimed function, namely, to analyze the user's voice and to classify it as a first or second voice." *Id.* (citing Pet., 11). As another example, the Petition describes the '337 patent's proximity sensor as "calculating the user-to-device distance, the distance can be used as a 'feature value' to perform the classification as the first or second voice." *Id.* Ground 1 does not address the details or arrangement of the components/steps of any claim of the '337 patent as required for anticipation. Petitioners tacitly admit that deficiency via introducing Ground 5. Petitioners attempt to add Kristjansson in Ground 5 to assert disclosure of certain recited components that perform certain functions. *Id.*, ¶83. But even then, Petitioners fail to address the independent claims. *Id.* They only attempt to identify a few of the recited components in dependent claim 2. *Id.* (citing Pet., 65-68). The Petition only applies Kristjansson to claim 2[a]. It skips claims 1 and 3-5. But even that attempt is cursory. The Petition spends most of its pages repeating incomplete obviousness rationales. Pet., 66-68.

In short, Petitioners admit, via Ground 5, of the necessity to identify the detailed arrangement of components/steps recited in the other claims. But they failed to do so in Ground 1.

In view of the foregoing, claims 1-5 are valid over Ground 1.

B. Ground 2: Shin Does Not Render the Challenged Claims Obvious

Ground 2 provides no basis on which to invalidate any claim beyond what claim 1 provides.

As a threshold matter, Petitioners' lead-in to every claim in Ground 2 is improper. Each is a cursory and generalized attempt at a catchall: "If PO argues that Shin fails to disclose *any* limitations, such limitations would have been obvious over Shin." Pet., 30 (emphasis added); *see also id.*, 32 (claim 2), and 33 (claim 3). That position is generalized. It is not developed. And it improperly conflates anticipation and obviousness. Obviousness and anticipation are two separate legal concepts with different requirements and elements.

The only obviousness ground (e.g., under the second *Graham* factor) actually disclosed in connection with any claim is based upon one inapplicable disclosure missing from Shin: the lack of disclosing two and only two classifications. No party asserts that the '337 patent claims require classification of two and only two voices. Thus, Ground 2, to the extent properly asserted, is moot. And Petitioners' attempt at a conclusory catchall should be disregarded.

Ground 2 also fails substantively. It is moot. But it is a microcosm of the problem with Petitioners' entire approach. Addressing disclosures of outputs does not necessarily provide information as to how inputs are processed—certainly

without having made an inherency argument. Petitioners' approach here is emblematic. Forcing a limitation on the outputs would not necessarily mean that Shin would have had to change anything about how inputs are processed. EX-2013, ¶84. Modifying Table 3 as Petitioners suggest would limit the amount of distance-based *outputs*. That is not what the classification limits of the claims recite. They recite classification of *input* voices.

Further, the Petition provides no actual reasoned analysis—required under *Graham* and *KSR* in Ground 2. In connection with claim 1, it provides six curt statements that partially resemble obviousness rationales. But when the boilerplate language is removed, the assertions are bare and conclusory—especially in terms of what was known and why the result would have been predictable. Ground 2 fails to provide even that much in terms of obviousness. In connection with claim 2, the Petition provides mere *ipse dixit*: “A POSITA would have recognized that implementing a[] [claim language] to [reorganized claim language], alongside a[] [claim language] to [restated claim language], would have been a predictable and logical extension of Shin's functionalities.” Pet., 32. That cannot possibly suffice under Supreme Court precedent. It would be legal error to grant the Petition on such grounds.

The Petition's analysis of claim 3 also provides curt, partial recitations of obviousness rationales. They are problematic for the same reason discussed in

connection with claim 1. But, even if they were complete, they do not address the *claim* in full context. All arguments are based solely on reducing words in an output (e.g., “[a]voiding repetition,” “reducing unnecessary repetition,” “enhance efficiency and conciseness”). Pet., 34-35. Claim 3 depends from claim 1. Claim 1 requires omission of information in connection with analysis and classification of voices based on distance. None of the purported rationales address the additional limitation in claim 3 in the context of those applicable limitations.

Finally, the Board recognized that the Petition provides no basis for modifying Shin in Ground 2. Dec., 18. Likewise, the Petition provides no grounds for any single-reference obviousness assertion beyond the moot and inapplicable grounds above (arguing for limiting a determination to two distance ranges).

C. Ground 3: Shimomura Alone Does Not Render the Challenged Claims Obvious

1. The Shimomura Translation Is Unreliable

Shimomura is a translation of a Japanese reference. This IPR turns on the precise disclosure of “voice,” “analysis,” and “classification” based upon characteristics those. There is no indication in the record that any particular consideration was given to the Japanese or English disclosures in Shimomura in that regard.

The Petition relies heavily upon purported expert testimony to address the lack of express disclosures in Shimomura. The lack of express disclosures elevates the potential for dispute over the context in which specific words are used. Such disputes go beyond whether a translation is certified as either true or not. They proceed into the realm of expert testimony. *See, e.g., Kurotaki v. United States*, No. 22-00063 MWJS-WRP, 2024 U.S. Dist. LEXIS 215953, at *13-15 (D. Haw. Nov. 27, 2024); *Hayashi v. Ozawa*, No. 17-cv-2558 (AJN), 2018 U.S. Dist. LEXIS 53559, at *5 (S.D.N.Y. Mar. 19, 2018). Ms. Snorteland has not laid the foundation for her understanding of the Japanese language (*e.g.*, fluency), let alone laid the foundation for her qualification as an expert on the distinction between “voice” and “distance” or “volume” in the context of the claimed “analy[sis] and “classif[ication].” *Kurotaki*, 2024 U.S. Dist. LEXIS 215953, at *15.

2. Shimomura Discloses Measurement of Distance, Not Classification of Voices

The Petition relies on Shimomura's measurement of distance to meet all limitations of the claims. It does not. As discussed in Section IV.B-C above, limitations 1[b], 4[c], and 5[c] require classes of *voices* (*i.e.*, “a first *voice* and a second *voice*”). EX-2013, ¶85 (citing EX-1001).

Shimomura does not disclose any manner of classifying voices. *Id.*, ¶86. The Petition acknowledges that Shimomura discloses a “device” that “controls speech

forms such as speech volume ... according to the *physical distance* to the conversation partner.” Pet., 36-37 (emphasis added) (citations omitted). The only basis the Petition provides for “classification” of a “voice” is in connection with Fig. 7. EX-2013, ¶86. The Petition does not identify any express disclosure of either aspect. *Id.* It does not cite to a single disclosure of a “voice.” *Id.* All mentions of “voice” (including those with quotes around them) are taken from Petitioners’ expert or the ’337 patent. *Id.* Neither the Petition nor Petitioners’ expert provides any reasoned analysis supporting the conclusion that Shin classifies voices. *Id.* Thus, regardless of the construction (e.g., the Board’s preliminary construction or the plain meaning), the Petition fails to demonstrate that Shimomura discloses a voice. *Id.*

The Petition also provides no reasoned analysis for why a POSITA would equate distance measurements with either the concept of a “voice” or “classification.” And it provides no inherency assertion.

Similarly, to the extent Shimomura discloses any required, claimed classification, it discloses classes of distances, not voices. *Id.*, ¶87. Petitioners argue that Shimomura discloses, in connection with Figure 7, “classifying voice into *two categories based on the user-to-device distance*: (1) a first voice *at greater than 350 cm*, and (2) a second voice *at less than 350 cm*.”

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

Less than 350 cm	350 cm or more
None	Implementation

Id. (citing Pet., 40; EX-1004, Fig. 7 (annotated by Petitioners)).

Petitioners' expert added the "First Voice" and "Second Voice" labels to *Id.*, ¶88 (citing Fig. 7. EX-1002, ¶64). Annotations aside, Figure 7 discloses one thing: a calculation of distance and nothing else. *Id.* That determination of distance cannot expressly disclose a classification because it is a measurement of a continuous variable, not discrete classification. *Id.* Nor can it expressly disclose an analysis of a "voice," which is not at all synonymous with "distance." *Id.* The multiple-users example described above in section VI.A.1 applies with equal force here. *Id.*

In fact, Shimomura confirms the distinction between "voice," "distance," and "volume." *Id.*, ¶89. It states the following:

However, conventional robots equipped with such *voice* recognition and dialogue control functions are constructed to always converse with the user at a constant speech volume set in advance, regardless of the physical *distance* between the robot and the user.

Id. (citing EX-1004, ¶[0004] (emphasis added)).

Thus, a POSITA would never have understood Shin's disclosure of distance to be synonymous with "voice." *Id.*, ¶90. Nor would a POSITA have understood Shin's determination of a distance to be synonymous with a classification. *Id.*

In short, like Shin, Shimomura directly determines the output content and volume based on distance without performing the step of classifying the voice. *Id.* ¶91 (citing EX-1004, ¶¶[0058], [0067]-[0070]). And applying Petitioners' expert's interpretation of voice classification, the input (distance) and output (device content) are likewise not voices. Thus, to the extent Shimomura discloses classification, it discloses classes of distance—not voices similar to as discussed for Shin. *Id.*

The above reasoning applies even under the Board's preliminary construction. *Id.*, ¶92. That is, even if a new voice could theoretically be created based solely on a volume change based on distance, that does not mean that a new voice is always created simply because distance is changed. *Id.* In other words, the concepts of "volume" and "distance" would still be broader than the concept of a "voice." *Id.* The Petition provides no basis upon which the Board could conclude that Shimomura discloses the concept of determining a voice as opposed to simply making a measurement of "distance." *Id.* And certainly a POSITA would never have understood Shin's disclosure of calculating a change in distance of a user,

who otherwise does not change their voice at all, as any sort of voice classification—even if volume changed as well. *Id.*

3. Shimomura Does Not Disclose Analysis of Voice for Classification

Limitation 1[b] recites: “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.*” *Id.*, ¶93 (citing EX-1001, Claim 1). Limitations 4[c] and 5[c] are similar for purposes of this discussion. *Id.*

As discussed in Section IV.1, those limitations require the claimed voice classification to be based on: **(1)** analysis of the acquired voice; and **(2)** distance separately calculated (*e.g.*, by a proximity sensor) in limitations 1[a], 4[a], and 5[a]. *Id.*, ¶94; *see also supra* §IV. None of the asserted grounds and prior art references therein discloses or suggests the claimed “analyz[e/ing] the [acquired] voice to classify the [acquired] voice.” *Id.*

Petitioners argue that Shimomura discloses “analyze a voice spoken by a user and acquired by a voice acquiring unit” allegedly because Shimomura’s “speech recognition unit 42 performs various sound-related recognition processes such as speech recognition, speaker recognition, and sound source direction recognition based on the speech signal S2 provided the microphone 21.” Pet., 37-

39 (quoting EX-1004 ¶[0040]). Petitioners argue that Shimomura discloses “analyze a voice ... to classify the voice ... based on the distance between the user and the voice-content control device” allegedly because Shimomura’s speech synthesis unit and speech character string transformation unit “transform the character string’ of the content spoken by the robot or change its volume ‘according to the distance to the conversation partner.’” *Id.*, 39 (quoting EX-1004 ¶¶[0058], [0067]-[0070]). Based on that, Petitioners allege that Shimomura discloses, in connection with Figure 7, “classifying voice into two categories based on the user-to-device distance: (1) a first voice at greater than 350 cm, and (2) a second voice at less than 350 cm.” *Id.* The Shimomura-based grounds are unavailing.

Like with Shin, Petitioners failed to allege that Shimomura’s determination of user-to-device distance (i.e., the basis of the alleged voice classification) involves any analysis of the user’s voice. EX-2013, ¶95 (citing Pet., 37-40). That is because Shimomura does not disclose determining user-to-device distance based on any voice analysis. *Id.* Even more, Amazon’s expert admitted to the same. EX-2015, 45:3-17, 53:17-22, 54:1-6. Rather, Shimomura relies exclusively on camera 20 in image recognition unit 41 to detect distance. EX-2013, ¶95 (citing EX-1004, ¶¶[0039] (“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 20”), [0059] (“... distance information indicating the distance to the conversation partner detected by the image recognition unit 41....”)).

Accordingly, Petitioners did not and could not demonstrate that Shimomura discloses “analyze a user’s voice ... to classify the voice” as recited in claim 1 (and similarly recited in claims 4-5) under both the plain meaning of the claims and Patent Owner’s construction because no voice analysis whatsoever is involved in Shimomura’s determination of user-to-device distance based by image recognition unit 41. *Id.*, ¶96.

4. Shimomura Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data

Claim 1 requires, *inter alia*, “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.” *Id.*, ¶97 (citing EX-1001, Claim 1). Limitations 4[h] and 5[h] are similar for purposes of this discussion. *Id.* The Board’s preliminary construction does not affect this limitation. *Id.*

Petitioners acknowledge that Shimomura “does not specify that the output sentences have different volumes.” *Id.*, ¶98 (citing Pet., 47). That is because

Shimomura does not expressly disclose a connection between the cited disclosures.

Id. The Petition does not provide any analysis supporting the assertion that the processes disclosed in connection with Figures 7 and 9 are compatible with one another or if or how a POSITA would have understood them to be used in conjunction with one another. *Net MoneyIN*, 545 F.3d at 1370-71; *Abiomed*, Paper 9, 8-12. As much is demonstrated by the fact that those figures disclose transforming the output content and changing volume at different distances:

[FIG. 7]

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

[FIG. 9]

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

52

FIG. 9 Volume Change Table

EX-2013, ¶¶98-99 (*comparing* EX-1004, Fig. 7 (transforming output content when user-to-device distance is 350cm or more) *with id.*, Fig. 9 (changing volume when user-to-device distance is less than 250cm)). Thus, in view of Figures 7 and 9 of Shimomura, Petitioners' alleged "first output sentence" (i.e., transformed character string at distance of 350 cm or more) and "second output sentence" (i.e., original character string at distance of less than 350 cm) would have the same volume at

distance more than 250 cm (as provided in Figure 9 of Shimomura). *Id.*, ¶¶99-100 (citing Pet., 43-46; EX-1004, Figs. 7, 9).

For that reason, Petitioners argue that “[a] POSITA would have been motivated to modify Shimomura’s volume change table to include a range ending at 350 cm,” either by “extending the 150-250 cm range to end at 350 cm,” as shown in Petitioners’ “First Modified Volume Change Table,” or by “adding a new range from 250 cm to 350 cm,” as shown in Petitioners’ “Second Modified Volume Change Table”:

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

First Modified Volume Change Table

Less than 50 cm	50~80 cm	80~150 cm	150~250 cm	250~350 cm	More than 350 cm
0	+1	+2	+3	+5	+7

Second Modified Volume Change Table

Id., ¶100 (citing Pet., 47-48).

But again, the Petition never establishes that Shimomura disclosed use of the tables in conjunction with one another. A POSITA, then, would have had no reason to modify the tables at all. They are tied to separate “processing procedures.” *See, e.g.*, EX-1004, Figs. 6 and 8; EX-2013, ¶¶ 101-102.

Even assuming they were part of the same process, a POSITA would not have had any reason to alter Shimomura's expressed teachings in the manner the Petition suggests. EX-2013, ¶101.

As Petitioners acknowledge, Shimomura expressly discloses transforming the output content and changing volume at different distances. *Id.*, ¶102 (citing EX-1004, Figs. 7, 9). *Id.* This is because Shimomura discloses that different options work at different distances. *Id.* It recognizes that “there is a certain *limit to the speech volume* of the robot 1,” such that “it *may not be possible* even if the *speech volume should be determined to be a larger volume in accordance with the distance* with the user when determining the volume in step SP22.” *Id.*, (citing EX-1004, ¶¶[0072]-[0073]). When the user is at a distance from Shimomura's robot where further increase in volume is not possible or not effective in making the robot's speech easier for the user to hear, Shimomura discloses changing the robot's speech form (other than speech volume), including speech speed, pause length, phrase-end emphasis, and most drastically, transforming the speech character string by adding “ne” to the end of each phrase. *Id.* (citing EX-1004, ¶¶[0060]-[0066], [0074]-[0079], [0084]-[0088]).

As shown in Figures 7, 10-12, Shimomura discloses changing the robot's speech form at user-to-device distances that do not align with those disclosed in Figure 9 for changing speech 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

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

[FIG. 10]

Less than 200 cm	200~350cm	350 cm or more
0%	+20%	+50%

53

FIG. 10 Speed change table

[FIG. 11]

Less than 100 cm	100~350cm	350 cm or more
0%	+30%	+60%

54

FIG. 11 Pause length change table

[FIG. 12]

Less than 200 cm	Less than 200 cm
None	Raise

55

FIG. 12 Phrase-end emphasis determination table

Id., ¶103 (citing EX-1004, Figs. 7, 9-12).

Accordingly, a POSITA would have understood Shimomura to teach that increasing speech volume alone may not be possible or sufficient for a robot's speech to be easily understandable by a user, especially at a greater user-to-device distance. *Id.*, ¶104 (citing EX-1004, ¶¶[0072]-[0073]). A POSITA also would have understood that, in addition to changing speech volume, other speech forms, such

as speech speed, pause length, phrase-end emphasis, and speech character string could also have been changed depending on user-to-device distance to make Shimomura's robot's speech easier to understand. *Id.* (citing EX-1004, ¶¶[0060]-[0066], [0074]-[0079], [0084]-[0088].). A POSITA further would have understood from Shimomura that the user-to-device distance at which the various speech forms discussed above should be changed would have differed from one speech form to another. *Id.* (citing EX-1004, Figs. 7, 9-12). For instance, in view of Figures 7 and 9-12 of Shimomura, a POSITA would have understood that at a user-to-device distance of less than 350 cm, a combination of changes to speech volume, speech speed, pause length, and phrase-end emphasis would have been sufficient for Shimomura's robot's speech to be understandable without transforming the speech character string. *Id.* (citing EX-1004, Figs. 7, 9-12). Logically, that makes sense. *Id.* At such closer distances, transforming the speech character string (e.g., by adding "ne" to the end of each phrase, as taught by Shimomura) would have been unnecessary, unnatural, and potentially annoying to a user. *Id.*

Thus, contrary to Petitioners' proffered hindsight modification of Shimomura, a POSITA would have recognized that Shimomura teaches transforming a speech character string (e.g., by adding "ne" to the end of each phrase, as taught by Shimomura) only when changes to other speech forms (e.g.,

speech volume, speech speed, pause length, phrase-end emphasis) are insufficient for the robot to be easily understandable by the user. *Id.*, ¶105. To the extent a POSITA would have been motivated to modify Shimomura's user-to-device distance at which to change the various speech forms, such a skilled person nonetheless would not have been motivated to set the distance at which to change speech volume to be the same as the distance at which to transform the speech character string. *Id.* That is because a POSITA would have recognized (in view of Shimomura's teachings in connection with Figures 7 and 9-12) that, even when Shimomura's robot output speech is at maximum volume, a combination of speech speed, pause length, and phrase-end emphasis could still have been changed. *Id.* Doing so would have improved the ease of comprehension of the robot's speech before it was necessary to transform the speech character string. *Id.* In fact, a POSITA would have understood Shimomura's choice to set different distances as discouraging a POSITA from changing them to be the same. *Id.*

The testimony of Petitioners' expert admits of the lack of any basis for Ground 3. He evidently disregarded all the foregoing disclosures, because he stated that he did not "know why" the numbers were different in the different tables. EX-2015, 101:15-102:19. He then abruptly concluded that it must have been due to "carelessness." *Id.* None of those provide the requisite legal basis for obviousness. There is no basis upon which to conclude that the drafter was careless. And one

cannot provide an expert conclusion that a POSITA would have been motivated to change numbers in a table in a certain way without knowing why a POSITA would have understood the purpose for the initial numbers.

5. Shimomura Does Not Disclose Generation of Output Based on Voice Classification

Claim 1[f]-[g] recite: “generate a first output sentence as the output sentence when *the acquired voice has been classified as the first voice*, and 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*.” EX-2013, ¶106 (citing EX-1001, Claim 1). Limitations 4[g] and 5[g] are similar for purposes of this discussion. *Id.* The Board’s preliminary construction does not affect this limitation. *Id.*

The Petition relies on Shimomura to assert that the claimed omitted information is disclosed in the form of “ne.” *Id.*, ¶107 (citing Pet., 46). But Shimomura states that such disclosure is a “change[] in intonation.” *Id.* (citing EX-1004, ¶[0092] (“changes in intonation (emphasis at the end of each phrase) ...”). It is not an omission of information. *Id.*

For reasons discussed in Sections VI.E.1-2 above, none of the asserted grounds and prior art references therein discloses or suggests classifying the acquired voice as the first and second voice as claimed. *Id.*, ¶108. For the same

reasons, none of the asserted grounds discloses or suggests generating a first and second output sentence when the acquired voice has been classified as the first and second voice. *Id.*

D. Ground 4: Shimomura in View of Shin Does Not Render the Challenged Claims Obvious

1. The Combination of Shimomura and Shin Discloses Measurement of Distance, Not Classification of Voices

The combination of Shimomura and Shin does not disclose limitations 1[b], 4[c], and 5[c] for the reasons discussed above in Sections VI.A.1 and VI.C.2.

2. The Combination of Shimomura and Shin Does Not Disclose Analysis of Voice for Classification

The combination of Shimomura and Shin does not disclose limitations 1[b], 4[c], and 5[c] for the reasons discussed above in Sections VI.A.2 and VI.C.3.

3. The Combination of Shimomura and Shin Does Not Disclose Generation of Output Based on Voice Classification

The combination of Shimomura and Shin does not disclose limitations 1[f]-[g], 4[g], and 5[g] for the reasons discussed above in Sections VI.A.3 and VI.C.5.

4. The Combination of Shimomura and Shin Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data

As discussed in Sections VI.A.4 and VI.C.4, neither Shin nor Shimomura alone would have rendered obvious the volume adjustment required by all the claims. EX-2013, ¶109.

For Ground 4, Petitioners assert that those limitations would have been obvious in view of Shin. *Id.*, ¶¶110-11 (citing Pet., 59-62). But nothing in Shin would have motivated a POSITA to deviate from Shimomura's expressed teachings. *Id.* At best, Shin suggests to a POSITA that the user-to-device distance at which to change the output volume and output content *could* be the same. *See Palo Alto*, 122 F.4th at 1384 (the "reason why" must be articulated); *In re Anova*, 809 F. App'x at 843. In the context of Shimomura, however, a POSITA would have recognized that doing so would have been undesirable as it would have burdened Shimomura's robot and user with superfluous phrases (e.g., "ne") when they are unnecessary for the robot to be understood. EX-2013, ¶111.

Tellingly, Petitioners' proffered motivation to combine is pure hindsight—Petitioners allege that their proposed modification of Shimomura in view of Shin "would enable Shimomura's device to *classify a voice into one of two distance-based categories* and adjust output volume level and content based on that category, rather than referencing tables with distinct ranges" (Pet., 61)—but Shin, like Shimomura, discloses tables with more than two distance ranges. *Id.*, ¶112 (citing EX-1003, Tables 1, 3 and Pet., 61). Neither Shin nor Shimomura discloses classifying a voice into one of two categories. *Id.* (citing Pet., 30-35 ("A POSITA would have been motivated for several reasons to implement Shin's volume change table (Table 1) and amount of information table (Table 3) with only two distance

ranges.”)). Thus, it is plain that Petitioners’ proffered motivation to combine Shimomura and Shin came not from the references themselves, but from the disclosures of the ’337 patent.

In short, for reasons discussed in Section IV.E.3 and above, a POSITA would not have been motivated by Shin to modify Shimomura in a manner that renders obvious the claimed volume adjustment. *Id.*, ¶113.

E. Ground 5: The Combination of Shin, Shimomura, and Kristjansson Does Not Render the Challenged Claims Obvious

1. The Combination of Shin, Shimomura, and Kristjansson Discloses Measurement of Distance, Not Classification of Voices

No combination of Shin, Shimomura, or Kristjansson discloses limitations 1[b], 4[c], and 5[c] for the reasons discussed above in Sections VI.A.1 and VI.C.2.

2. The Combination of Shin, Shimomura, and Kristjansson Does Not Disclose Analysis of Voice for Classification

No combination of Shin, Shimomura, or Kristjansson discloses limitations 1[b], 4[c], and 5[c] for the reasons discussed above in Sections VI.A.2 and VI.C.3.

3. The Combination of Shin, Shimomura, and Kristjansson Does Not Disclose Generation of Output Based on Voice Classification

No combination of Shin, Shimomura, or Kristjansson discloses limitations 1[f]-[g], 4[g], and 5[g] for the reasons discussed above in Sections VI.A.3 and VI.C.5.

4. The Combination of Shin, Shimomura, and Kristjansson Does Not Disclose the Claimed Adjusting of the Sound Volume of Voice Data

No combination of Shin, Shimomura, or Kristjansson discloses the claimed sound adjustment for the reasons discussed above in Section VI.D.4.

F. Grounds 2-5: The Petition's Obviousness Analysis is Legally Deficient

1. Illogical Modifications

The tables the Petition cites in Shin and Shimomura are concerned only with making a discrete measurement: of distance between a device and a user. Each would operate exactly the same regardless of whether any voice changed, so long as the distance remained constant. EX-2013, ¶114. The simplicity of that metric demonstrates that the authors of Shin and Shimomura specifically declined to conduct an analysis of voices to classify voices, let alone based on distance. *Id.* They purposely chose simpler, more discrete approaches. *Id.* Each obviousness position thus fails as a matter of logic. Each purports that the alleged modification or combination would have made the devices simpler and more efficient. But adding the claimed analysis and voice-classification would have achieved the opposite. *Id.* It would have mandated a more complex approach than implementing a single-variable measurement in simple, two-dimensional tables. *Id.*

2. Failure to *Apply* Second Graham Factor

“[I]t is not enough for [a challenger] to merely demonstrate that elements of the claimed invention were independently known in the prior art.” *Metalcraft of Mayville, Inc. v. Toro Co.*, 848 F.3d 1358, 1367 (Fed. Cir. 2017) (citation omitted).

“Often, every element of a claimed invention can be found in the prior art.” *Id.*

Additionally, merely identifying a problem and a potential solution is not enough.

“[W]e cannot allow hindsight bias to be the thread that stitches together prior art patches into something that is the claimed invention.” *Id.* (citing *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 419, 421 (2007)). That is where *Graham* comes in.

Petitioners fail to apply *Graham* factor two in their obviousness analysis for Grounds 2-5. The Petition does not identify “the differences between the claims and the prior art.” *Arctic Cat Inc. v. Bombardier Rec. Prods.*, 876 F.3d 1350, 1358 (Fed. Cir. 2017) (citing *Graham v. John Deere Co.*, 383 U.S. 1, 17-18 (1966)).

Petitioners must address all four *Graham* factors in their obviousness analysis.

WBIP, LLC v. Kohler Co., 829 F.3d 1317, 1328 (Fed. Cir. 2016) (“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 those factors are considered.” (citations omitted)). *Graham* factors one and two go hand-in-hand, and motivation to combine and expectation of success are generally considered as part of the first factor. *Dystar Textilfarben GmbH v. C.H.*

Patrick Co., 464 F.3d 1356, 1370 (Fed. Cir. 2006); *Apple Inc. v. Samsung Elecs. Co., Ltd.*, 839 F.3d 1034, 1061 (Fed. Cir. 2016); *Pfizer Inc. v. Sanofi Pasteur Inc.*, 94 F.4th 1341, 1348 (Fed. Cir. 2024); MPEP § 2143.

Petitioners do not explain what specifically is missing from the primary references or why a POSITA would have been motivated to bridge that particular gap. Petitioners thus do not sufficiently address *Graham* factor two at all.

That is evident from its expert's deposition. He was unable to articulate the *differences* between Shin and claim 1 of the '337 patent beyond limiting the claim to a single threshold value. EX-2015, 109:12-112:16. In connection with Shimomura, he was only able to identify a difference in values in tables between Shimomura and *Shin*—not between either reference and claim 1. *Id.*, 113:3-117:5. Neither approach satisfies *Graham* factor two—especially under *KSR*.

The Petition only attempts to address *Graham* factor two in limited and isolated instances. For example, the Petition raises six different potential combinations in Grounds 1-5. And it proposes obviousness grounds to invalidate every limitation of every claim. And yet the following are the only instances in which the Petition even attempts to provide a *Graham* factor two analysis.

In Ground 2, the Petition states “[e]ven if PO were to argue that claim 1 requires the user’s voice to be classified into no more than two classifications based on distance, Shin renders this obvious.” Pet., 30.

In Ground 3, Petitioners state that “[a]lthough Shimomura generally discloses altering the volume based on distance, it does not specify that the *output sentences have different volumes*.” *Id.*, 47 (emphasis added).

In Ground 4, Petitioners state that “Even if Shimomura alone did not teach or motivate a POSITA to *alter the output’s content and volume at the same threshold distance . . .*” *Id.*, 59 (emphasis added).

In Ground 5, Petitioners state that “[s]pecifically, to the extent PO argues that Shin and/or Shimomura do not disclose the “*process executing unit*” limitations (*e.g.*, claim limitations 1[c], 2[a], 4[d], and 5[d]), Kristjansson discloses those limitations and renders the claims obvious.” *Id.*, 65 (emphasis added).

Petitioners make no obviousness argument other than modifying Shin to have two and only two classifications. As described above, Ground 4 at most identifies a difference between two references—not addressing the claims.

Patent Owner requested that it be provided notice of what these positions are. POPR, 34. That is, each applicable difference between each claim and the prior art. Only then would Patent Owner have been provided notice of how Petitioners apply *Graham* factor 2 (*i.e.*, the specific differences for which secondary references are added) in its prima facie case. Neither the Board nor the Petition has done so. Thus, for this additional reason, the IPR cannot succeed, as a matter of notice and constitutional law.

3. No Reason for the Combinations

Petitioners indiscriminately select disclosures from the various references to attempt to piece together the claim limitations. The Petition never addresses why a POSITA would have seen any reason to combine or modify the references to fill in those gaps. In contrast, doing so would defeat the core purpose of the disclosures. Petitioners do not address those problems because the Petition's obviousness analysis is too superficial to do so. Nor does the Petition ever analyze the claims as a whole. It simply focuses on discrete limitations. When addressing Shin and Shimomura in combination, for instance, the Petition zooms in on two tables without any discussion of the larger purpose or context of any claimed invention. *See* §VI.D. *supra* (discussing improper hindsight).

Petitioners never identify either (1) what problem a POSITA would have sought to solve when combining references or (2) any non-redundant benefit disclosed in them. *Id.* An inquiring party might ask why Petitioners assert that a POSITA would have combined Shin and Shimomura and possibly Kristjansson. There is no answer in the Petition—as discussed below.

4. Obviousness Arguments Are Perfunctory

The Petition does not establish that any results would have been predictable or any reasonable expectations of success. In many instances, the Petition merely recites the words “predictable” or “reasonable expectation” without providing any

support or explaining how or why these results would have had any reasonable expectation of success. *See, e.g.*, Pet., 31, 32, 34, 48, 49, 50, 57, 58, 60, 61, 63, 64, 67. Those sorts of statements lack support and are legally insufficient. *Broadcom Corp. v. Emulex Corp.*, 732 F.3d 1325, 1335 (Fed. Cir. 2013) (“Even assuming that a person of ordinary skill might have some motivation to [combine], the record does not show any reasonable expectation that this significant change would be successful.”); *A/S v. Coolit Sys. Inc.*, No. 19-cv-00410-EMC, 2022 U.S. Dist. LEXIS 246288 at *23-24 (N.D. Cal. Sep. 11, 2022) (the following “conclusory sentence” insufficient: “[w]hen the cited references are combined or modified as a POS[IT]A would have been motivated to do, the combination or modification is also based on conventional or known methods that would have yielded predictable results and been reasonably expected to be successful by a POS[IT]A.’ ... This conclusory assertion is insufficient show reasonable expectation of success.”).

5. The Federal Circuit Has Rejected the Proffered Motivations to Combine

All purported motivations to combine in the Petition have been rejected. They are purely generic and bear no relation to the specific claims. Each one accomplishes at most two things: (1) stating a motivation to improve a device and (2) asserting that the combination of prior art would have done so. Neither supports any element of a proper obviousness analysis. Every POSITA would have sought

to improve every device. But merely repeating that desire fails to state *why* a POSITA would have sought the *particular* solution (combination) proposed over the innumerable other possibilities. It does not state why a POSITA would have chosen one route over any other. Thus the missing piece—*why* a POSITA would have chosen the *particular* solution (combination)—is pure hindsight.

They Petition provides the following motivations:

- “simplifying complex systems into fewer categories . . .” Pet., 31.
- “to simplify user experience.” *Id.*, 31.
- “simplifying implementation and minimizing computational demand.” *Id.*, 32.
- “simplifies design by minimizing programming complexity.” *Id.*, 49.
- “simplifying design . . .” *Id.*, 61.
- “perform these functions more efficiently” *Id.*, 55.
- “an efficient, user-friendly solution” *Id.*, 31.
- “a streamlined, efficient, and straightforward user interface” *Id.*, 31.
- “efficient, streamlined interactions” *Id.*, 34.
- “to configure the processor for more efficient and streamlined interactions.” *Id.*, 35.
- “perform these functions more efficiently” *Id.*, 55.
- “excessive repetition would be inefficient” *Id.*, 57.

- “efficient intent extraction and content retrieval” *Id.*, 62.

Each is more generic, and less specific, than the following:

The motivation to combine would be because you wanted to build something better. You wanted a system that was more *efficient*, cheaper, or you wanted a system that had *more features*, makes it more attractive to your customers, because by combining these two things you could do something new that hadn't been able to do before.

ActiveVideo Networks, Inc. v. Verizon Communs., Inc., 694 F.3d 1312, 1328 (Fed. Cir. 2012) (emphasis added). The Federal Circuit rejected that genericism as “bear[ing] no relation to any specific combination of prior art elements.” *Id.* (citing *KSR*, 550 U.S. at 418).

Making systems simpler and more efficient are two related, ubiquitous concepts. They lack any particular meaning in an obviousness analysis. *Hulu, LLC v. Sound View Innovations, LLC*, IPR2018-00582, Paper 34, 21-22 (PTAB Aug. 5, 2019) (informative) (“a generic desire for ‘improved access times’ is “merely a bald statement ...”). “The factual inquiry whether to combine references must be thorough and searching, and the need for specificity pervades this Court’s authority on the Board’s findings on motivation to combine.” *Id.* (citation modified).

6. Improper Circumvention of Obviousness and Anticipation

Grounds 4 and 5 appear to be attempts to avoid the proper obviousness and inherency standards. Petitioners disclose those grounds in quick, cursory fashion. Pet., 59-67. The large majority of the applicable pages in the Petition is spent repeating around eight incomplete versions of the obviousness rationales from MPEP § 2143.

Petitioners asserted that a POSITA would have combined the references for the purpose of outputting content and volume at a single, *numerical* threshold distance. The Petition does not explain, however, how that combination/modification relates to any claim limitation. Instead, the Petition attempts to create an artificial steppingstone between assessing the prior art and the claims. That approach undermines established precedent against hindsight reasoning, using the invention as a roadmap to piece together the invention.

7. Failure to Provide Fair Notice

The Petition and expert declaration each repeatedly state contradictory alternative positions without differentiating between the two: “known and obvious,” “anticipate and render obvious,” “disclose and render obvious,” “disclose or render obvious.” *See, e.g.*, Pet., 59, 62, 64-65, 69; EX-1002, ¶¶175, 186, 198, 216, 217 (*e.g.*, “Shin and Shimomura disclose and/or render obvious each limitation”). That is confusing, non-committal, multiplicative language.

8. Expert Declaration Inseparable from Attorney-Argument

The Petition relies solely on indeterminate statements to address important claim limitations (*e.g.*, voice classification and operations based on it). The language in the Petition mirrors statements in the expert declaration in that regard. So, it is not clear whether the statements are intended to constitute expert opinion, legal argument, or something else. *Compare* EX-1002, ¶¶117-120 *with* Pet., 32-35 (Ground 2); *Compare* EX-1002, ¶¶129-197 *with* Pet., 35-59 (Ground 3); *Compare* EX-1002, ¶¶199-204 *with* Pet., 59-64 (Ground 4); *Compare* EX-1002, ¶¶216-223 *with* Pet., 65-68 (Ground 5).

Petitioners' expert declaration should thus be given little or no weight. *See Xerox Corp. v. Bytemark, Inc.*, IPR2022-00624, Paper 9, 15-17 (PTAB Aug. 24, 2022) (precedential); *Facebook, Inc. v. Windy City Innovations, LLC*, 973 F.3d 1321, 1340–41 (Fed. Cir. 2020).

9. Modifying Outputs Does Not Disclose Classification of the Claimed Inputs

Ground 4 states that “[e]ven if Shimomura alone did not teach or motivate a POSITA to alter the output’s content and volume at the same threshold distance (*e.g.*, 350 cm), doing so would have been obvious in view of Shin.” Pet., 59. Ground 5 builds on that concept but cites Kristjansson regarding processing components. *Id.*, 65.

Petitioners argue that a “POSITA would have naturally combined these teachings to enhance adaptability and functionality in voice-controlled devices with predictable results.” *Id.*, 60. That statement is entirely based on the expert declaration which, as explained above, mirrors the Petition.

It is also inapposite. Adjusting the outputs does not address the missing limitation of classifying inputs (voices).

VII. CONCLUSION

For all of the foregoing reasons, the Petition fails to demonstrate that any one of challenged Claims 1-5 of the '337 patent is unpatentable.

Dated: December 9, 2025

Respectfully submitted,

By: /Steven J. Reynolds/

Steven J. Reynolds (USPTO Reg. No. 61,445)*

Kevin H. Sprenger (USPTO Reg. No. 79,466)

Chandran B. Iyer (USPTO Reg. No. 48,434)

Erin Hadi (USPTO Reg. No. 79,904)*

Louay Meroueh (USPTO Reg. No. 81,304)*

sreynolds@daignaultiyer.com

ksprenger@daignaultiyer.com

cbiyer@daignaultiyer.com

ehadi@daignaultiyer.com

lmeroueh@daignaultiyer.com

DAIGNAULT IYER LLP

8229 Boone Boulevard

Suite 450

Vienna, VA 22182

**Not admitted in Virginia*

Attorneys for SoundClear Technologies LLC

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that this Patent Owner's Response complies with the type-volume limitations of 37 C.F.R. §§ 42.24(b) because it contains 13,994 words, as determined by the word-processing program used to prepare the brief, excluding the parts of the brief exempted by 37 C.F.R. § 42.24(a)(1).

Dated: December 9, 2025

Respectfully Submitted,

By: /Steven J. Reynolds/

Steven J. Reynolds
USPTO Reg. No. 61,445

CERTIFICATE OF SERVICE

The undersigned hereby certifies that the foregoing PATENT OWNER'S RESPONSE was served electronically via email on December 9, 2025, on the following counsel of record for Petitioner:

Colin B. Heideman
Jeremy A. Anapol
Christie R.W. Matthaei
Nathan D. Reeves
Logan P. Young
MORGAN, LEWIS & BOCKIUS LLP
colin.heideman@morganlewis.com
jeremy.anapol@morganlewis.com
christie.matthaei@morganlewis.com
nathan.reeves@morganlewis.com
logan.young@morganlewis.com
amazon-soundclear@morganlewis.com

Dated: December 9, 2025

Respectfully Submitted,

By: /Steven J. Reynolds/

Steven J. Reynolds
USPTO Reg. No. 61,445