

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

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

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner.

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

**PATENT OWNER'S RESPONSE TO PETITION FOR
INTER PARTES REVIEW OF UNITED STATES PATENT NO. 11,244,675**

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
II. LEVEL OF ORDINARY SKILL IN THE ART	2
III. THE ’675 PATENT	3
A. Specification	3
B. Prosecution History	3
IV. CLAIM CONSTRUCTION	4
A. “Classify the Voice into Either a First Voice or a Second Voice”	5
B. “Whisper”	7
C. Inconsistent Claim Construction Positions	8
V. OVERVIEW OF THE ALLEGED PRIOR ART	9
A. Shin (EX-1003)	9
B. Shimomura (EX-1004)	9
C. Aoyama (EX-1010)	9
D. Iwase (EX-1017)	11
E. Schuster (EX-1016)	11
F. Kristjansson (EX-1013)	12
VI. THE CHALLENGED CLAIMS OF THE ’675 PATENT ARE VALID	12
A. Deficiencies Present in All Grounds	12
1. Equating Measurement to the Object Being Measured	12
2. Failure to Address Obviousness Under Prevailing Precedent ..	16

Patent Owner's Response

3.	Reliance on Hindsight.....	19
4.	Dr. Stern's Testimony Is Unreliable	21
B.	GROUND 1A: Shin in View of Aoyama Does Not Render the Challenged Claims Obvious.....	23
1.	Claims 1-7: Shin Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice	23
2.	Claims 1-7: Shin in View of Aoyama Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice	28
3.	Claim 5: Shin Does Not Disclose Determining the Voice Spoken by the User is a Whisper	33
4.	Claims 1-7: Shin Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice.....	37
5.	Claims 1-7: Shin Does Not Disclose the Claimed Components Arranged as in the Claims.....	38
6.	Claims 2-5	38
C.	GROUND 1B: Shin in View of Iwase Does Not Render the Challenged Claims Obvious.....	39
1.	Claims 1-7: Shin Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice	39
2.	Claims 1-7: Shin in View of Iwase Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice	39
3.	Claim 5: Shin Does Not Disclose Determining the Voice Spoken by the User is a Whisper	42
4.	Claims 1-7: Shin Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice.....	42

5.	Claims 1-7: Shin Does Not Disclose the Claimed Components Arranged as in the Claims.....	42
6.	Claims 2-5	43
D.	GROUND 2A: Shimomura in View of Aoyama Does Not Render the Challenged Claims Obvious.....	43
1.	Claims 1-7: Shimomura Discloses Measurement of Distance, Not Classification of Voices	43
2.	Claims 1-7: Shimomura in View of Aoyama Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice	46
3.	Claim 5: Shimomura Does Not Disclose Determining the Voice Spoken by the User is a Whisper	48
4.	Claims 1-7: Shimomura Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice.....	51
5.	Claims 1-7: Shimomura Does Not Disclose the Claimed Components Arranged as in the Claims	52
6.	Claims 2-5	52
E.	GROUND 2B: Shimomura in view of Iwase Does Not Render the Challenged Claims Obvious.....	53
1.	Claims 1-7: Shimomura Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice.....	53
2.	Claims 1-7: Shimomura in View of Iwase Does Not Disclose Replacing Words in a First Output Sentence When the Voice Is Determined to be the First Voice	53
3.	Claim 5: Shimomura Does Not Disclose Determining the Voice Spoken by the User is a Whisper	55

Patent Owner's Response

4.	Claims 1-7: Shimomura Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice	55
5.	Claims 1-7: Shimomura Does Not Disclose the Claimed Components Arranged as in the Claims	56
6.	Claims 2-5	56
F.	GROUND 3: Claims 3-4	56
G.	GROUND 4: Claim 5 is Not Obvious Over Grounds 1A-2B and Schuster	56
H.	GROUND 5: The Combination of Grounds 1-4 and Kristjansson Does Not Render the Challenged Claims Obvious	59
1.	Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice.....	59
2.	Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice	59
3.	Claim 5: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Determining the Voice Spoken by the User is a Whisper.....	60
4.	Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice	60
5.	Claims 2-5	60
I.	Other Deficiencies.....	60
1.	Illogical Modifications.....	60
2.	Failure to <i>Apply</i> Second Graham Factor.....	61
3.	Improperly Articulated Combinations for Ground 5	62

VII. CONCLUSION.....	63
----------------------	----

TABLE OF AUTHORITIES

Cases	Page(s)
<i>Allergan, Inc. v. Apotex, Inc.</i> , 754 F.3d 952 (Fed. Cir. 2014)	15
<i>In re Anova Hearing Labs, Inc.</i> , 809 F. App'x 840 (Fed. Cir. 2020)	2
<i>Arctic Cat Inc. v. Bombardier Rec. Prods.</i> , 876 F.3d 1350 (Fed. Cir. 2017)	61
<i>Jones v. Hardy</i> , 727 F.2d 1524 (Fed. Cir. 1984)	15
<i>In re Kahn</i> , 441 F.3d 977 (Fed. Cir. 2006)	21
<i>Kinetic Concepts, Inc. v. Smith and Nephew, Inc.</i> , 688 F. 3d. 1342 (Fed. Cir. 2012)	20
<i>KSR Int'l Co. v. Teleflex Inc.</i> , 550 U.S. (2007).....	17, 18, 19, 62
<i>Microsoft Corp. v. Enfish</i> , LLC, 662 F. App'x 981 (Fed. Cir. 2016)	18
<i>In re Omeprazole Patent Litigation</i> , 536 F.3d 1361 (Fed. Cir. 2008)	19, 20
<i>Palo Alto Networks, Inc. v. Centripetal Networks, LLC</i> , 122 F.4th 1378 (Fed. Cir. 2024)	2, 21, 41, 55
<i>Princeton Biochemicals, Inc. v. Coulter, Inc.</i> , 411 F.3d 1332 (Fed. Cir. 2005)	15
<i>Randall Mfg. v. Rea</i> , 733 F.3d 1355 (Fed. Cir. 2013)	35
<i>Ruiz v. A.B. Chance Co.</i> , 357 F.3d 1270 (Fed. Cir. 2004)	15

<i>Unigene Labs., Inc. v. Apotex, Inc.</i> , 655 F.3d 1352 (Fed. Cir. 2011)	17
<i>WBIP, LLC v. Kohler Co.</i> , 829 F.3d 1317 (Fed. Cir. 2016)	61

PATENT OWNER'S EXHIBIT LIST

Exhibit No.	Description
2001	Docket Printout of <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2002	Order from Judge Anthony Trenga, dated January 10, 2025 (Dkt. 84), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2003	Scheduling Order, dated September 23, 2024 (Dkt. 34), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2004	Final Scheduling Order, dated October 18, 2024 (Dkt. 55), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2005	Judge Anthony Trenga's median time to trial, available at https://law.lexmachina.com/shared/eyJzaGFyZWRfcGFnZV9pZCI6MTI3OTMxfQ.aA_aRw.A2n-SunBPfmUP2aka341_-3f27g
2006	Excerpts from the Motion to Stay hearing transcript, dated March 26, 2025, from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2007	Joint proposed scheduling order (Dkt. 48), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
<i>Exhibit Numbers 2008-2009 not used</i>	
2010	May, L., The Evolution of AI in Language Processing and its Impact on IP/Patents; May 9, 2025
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/MedianDispositionTimeMerits.pdf .

2012	U.S. Patent No. 11,069,337
2013	Motion to Dismiss, dated September 13, 2024 (Dkt. 32), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2014	11,244,675 Prosecution History
2015	Defendant's Invalidity Contentions, from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2016	Declaration of Patent Owner Expert Dr. Joseph C. McAlexander
2017	Transcript of Deposition of Petitioners' Expert Dr. Richard Stern
2018	Plaintiff SoundClear's Response Brief in Opposition to Defendants' Motion to Dismiss Under Rule 12(b)(6) (Dkt. 35), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2019	Plaintiff SoundClear's Motion Under Rule 59(e) To Alter or Amend Judgment and Order Granting Motion to Dismiss (Dkt. 65), from <i>SoundClear Techs. LLC v. Amazon.com, Inc. et al.</i> , CA No. 1:24-cv-01283 (EDVA)
2020	Transcript of Deposition of Petitioner's Expert Dr. Richard Stern in IPR2025-00565

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. 11,244,675 (“the ’675 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 1-7 of the ’675 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. A panel addressing a patent by the same inventor recognized that similar claim language—insofar as “voice” and “classification” is concerned—cannot possibly support the position Petitioners propose here. It determined, *sua sponte*, that the claims would have to be rewritten in order for the Petition to succeed. But neither party, nor the Board, has proposed any such construction here. Thus, the Petition fails for that reason alone.

Regardless, any such interpretation or application required by the Petition would constitute clear legal error. It would directly contradict the plain meaning of the core tenets of the claims, including voice classification.

That said, even under the construction the other panel proposed, none of the prior art discloses the full context of the claims. None demonstrate a classification

of voices, as opposed to merely measuring distance or volume. And none do 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) (citation omitted) (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 for its finding.”); *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 ’675 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-2016, ¶34.

III. THE '675 PATENT

The '675 patent is based on U.S. Patent Application No. 16/295,034, which was filed on March 7, 2019. EX-1001, 1. The '675 patent also claims priority to Japanese Patent Application No. JP2018-044598, which was filed on March 12, 2018. *Id.* Consequently, the priority date of the '675 patent is at least as early as March 12, 2018. EX-2016, ¶¶35-36.

A. Specification

The '675 patent discloses an “output-content control device” that includes “a voice classifying unit configured to analyze a voice spoken by a user.” EX-1001, Abstract; EX-2016, ¶¶37-40. The focus is on user interaction with a device by speaking to it. *Id.* More particularly, it relates to a device that analyzes a user's voice. That voice includes multiple aspects, including the “content” contained in it. *Id.* The '675 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.*

B. Prosecution History

During prosecution, the Examiner considered both Aoyama and Schuster as secondary references in combination with other prior art references. The Petition

relies on Aoyama (Grounds 1A, 2A, 3, 4, and 5) and Schuster (Grounds 4 and 5) as secondary references. Pet., 5; EX-2016, ¶¶41-42.

The Examiner cited both Aoyama and Schuster as secondary references in combination with other prior art references in a Non-Final Office Action and then in a Final Office Action. EX-2014, 127-135, 167-183, 211-226. The Examiner's extensive discussion of Aoyama and Schuster throughout prosecution shows that both references were properly considered.

IV. CLAIM CONSTRUCTION

No claim terms were proposed for construction in the Petition or by the Board in its Institution Decision ("Dec."). Patent Owner agrees that no claim construction is necessary. EX-2016, ¶43. That said, Amazon's arguments are founded on an interpretation that is far different than the plain and ordinary meaning of the claims. As much is demonstrated by the fact that the panel in the IPR2025-00565 ("-565") proceeding determined Amazon's position could only succeed if the claims were to be construed. So, it issued a preliminary construction. The '675 patent at issue here and the '337 patent at issue in the -565 proceeding are not technically related in a family. But they recite similar subject matter, including the classification of voices. The -565 panel's preliminary construction is problematic for the multiple reasons discussed in connection with that petition. But much of the same reasoning applies here. Amazon's Petition in this proceeding

cannot succeed without vitiating the plain and ordinary meanings of at least two claim phrases, within the context of the claims.

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

Independent claims 1, 6, and 7 each recite “calculat[e/ing] a distance between the user and [the/an] output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance.” EX-1001, claims 1, 6-7. That phrase, according to its plain and ordinary meaning, requires a classification of *voices*. It is not, and cannot be, met by continuous measurements of distance, or distance ranges. The plain language requires a delineation between “a first *voice*” and “a second *voice*.”

The specification makes clear that the claimed “first voice” and “second voice” are different classes of *voices* (e.g., a whisper voice, i.e., “an unvoiced voice spoken without vibrating the vocal chords”). *Id.*, 10:4-8. And the specification’s description of “feature values” (including user-to-device distance) and threshold thereof are determined and set for the purpose of distinguishing one class of voice from another. *See* EX-2020, 36:7-37:6; *id.*, 10:32-38 (“[T]he distance can be used as a feature value to perform the classification to the first voice V1A and the second voice V1B.”), 10:47-53 (“As described above, ... the voice classifying unit 38 classifies the voice V1 to the second voice V1B when the voice V1 is a whisper,

and to the first voice V1A when the voice V1 is not a whisper even when meaning of the voice V1 is the same.”).

Dr. Stern agrees that the concept of “voice” cannot be reduced to a simple metric like distance. EX-2020, 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. *Id.*, 28:12-18, 31:11-32:6.

To the extent the claimed voice classification is based on a single “feature value,” a POSITA would have recognized that the threshold thereof must be sufficient to delineate one class of voice from another. EX-2016, ¶¶44-46. For example, where volume is used as a “feature value” to classify a voice as either a whisper or not a whisper, a POSITA would have understood that such a volume threshold must be set low enough to distinguish a whisper from normal speech with reasonable accuracy. *Id.* A volume threshold that is consistent with both whisper and normal speech would not be fit for that purpose. *Id.* Similarly, where user-to-device distance is used as a “feature value,” such a distance threshold must be sufficient for the purpose of classifying voices. And as Amazon’s expert admitted during deposition, while distance may sometimes be sufficient to classify voices, that is not always the case. *See* EX-2020, 48:19-49:7.

Thus, whatever metric or metrics are used satisfy the claimed “classify the voice into either a first voice or a second voice based on the calculated distance”

only if the end result is to distinguish a first class of voice from a second class of voice. While distance can be associated with a voice, distance is not a voice attribute. Distance is not an aspect of a voice. Arbitrary distance measurements cannot alone meet that limitation. EX-2016, ¶47.

B. “Whisper”

SoundClear stated in its POPR that it would not seek a construction because Petitioners did not seek one. However, Amazon's position as to claim 5 relies on a de facto construction of whisper to simply mean a “range of speech volumes and rates.” Pet., 38; EX-2016, ¶48. There is no basis upon which to conclude that a POSITA would have understood “whisper” to relate to a “range of speech volumes and rates.” Dr. Stern confirmed as much:

A No. Shifting into whisper mode is *something other than* simply reducing the vocal volume.

See EX-2017, 86:4-13 (emphasis added); *see also* 89:15-90:13 (“I believe that Shin does not.”)); EX-2016, ¶49.

The '675 patent explains that a “whisper” results from an “unvoiced voice spoken without vibrating the vocal chords” EX-1001, 10:4-5. The patent indicates that the classification of a whisper does not have to be perfect: “[I]t is not limited to the unvoiced voice spoken without vibrating the vocal chords as long as being classified to the first voice V1A by, for example, a method described later.”

Id., 10:5-8. The method (identified below) related to whisper-determination follows: “For example, the voice classifying unit 42 determines the voice as a whisper to classify to the first voice V1A when the peak frequency is equal to or lower than the threshold, and determines the voice as not a whisper to classify to the second voice V1B when the peak frequency is larger than the threshold.” *Id.*, 10:17-22. The patent provides no other disclosure specifically related to determining a “whisper.” The patent discloses other methods of classifying “voice” but not in connection with determining a “whisper.” Thus, a POSITA would have understood that volume detection alone cannot suffice for determination of a whisper. EX-2016, ¶50.

In short, the Board should reject any argument from Petitioners that is based upon a whisper being equated to volume or speech rate.

C. 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-2016, ¶¶51-52 (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., 12 (emphasis added)).

B. Shimomura (EX-1004)

Shimomura is concerned only with distance and the volume output based on distance. EX-2016, ¶¶53-55. It thus lacks all voice classification limitations, especially those based on an analysis of the voice and distance. *Id.* (citing EX-1004, 1, ¶[0012]). Shimomura also discloses a “change[] in intonation.” *Id.* (citing EX-1004, ¶[0092]).

C. Aoyama (EX-1010)

Aoyama discloses “an information processing” device and method. EX-1010, Title.

It then discloses two methods of producing outputs. First, it discloses, among other things, a “response parameter generation unit 16.” *Id.*, ¶[0116]. The response parameter generation unit assesses information based on “natural

language processing on the text information obtained by converting the sound input from [a] speech style detection unit 115” *Id.* It also “acquires information indicating the degree of noise ... from [an] external environment estimation unit” *Id.* And it “acquires information indicating the user state or information regarding the user, such as the ‘degree of composure,’ the ‘age and the sex of the user,’ and the ‘recognition result of the individual user’ from [a] user state estimation unit 15.” *Id.* Aoyama further discloses a “response generation unit 17” that generates a response based on information from the response parameter generation unit. *Id.*, ¶¶[0123]-[0124]. The response generation includes inputting words based upon information from the response parameter generation unit.

Second, Aoyama separately discloses “control[ling] an amount of information output as the response” *Id.*, ¶[0145]. It does so based on “acquired information indicating the ‘degree of composure.’” *Id.* The “‘degree of composure’ ... is a parameter for determining a psychological state of the user (specifically, whether the user is calm).” *Id.*, ¶[0102].

The USPTO issued the ’675 patent over Aoyama. During prosecution, the Examiner considered Aoyama and specifically allowed the challenged claims over Aoyama. EX-2014, 127-135, 167-183, 211-226; EX-2016, ¶¶56-60.

D. Iwase (EX-1017)

Iwase discloses “an information processing device including a control unit” that modifies the position of a subject (e.g., a word) in a sentence that forms part of audible notifications provided to a user. EX-1017, Abstract, 3:24-29. Iwase also discloses two processes for generating outputs.

First, Iwase “determines that an attention acquisition difficulty level related to the user U1*a* is high, and therefore the information processing device may set a subject output position to the latter half of notification contentions.” *Id.*, 4:30-36. “[F]or example, in a situation in which a noise level is low, the information processing device determines that an attention acquisition difficulty level related to the user U1*b* is low, and therefore the information processing device may set a subject output position to the first half of notification contents.” *Id.*, 4:36-41.

Second, Iwase separately discloses a “speaking control unit 360 [that] shapes the extracted subject on the basis of attribute information of a user...” *Id.*, 15:6-10; EX-2016, ¶¶61-64.

E. Schuster (EX-1016)

Schuster discloses a “portable electronic communication device” that “determine[s] based on the user input whether the user desires the device to enter a private mode.” EX-1016, Abstract. Schuster operates by “mov[ing]” between a “normal mode” and a “private mode.” *Id.*, 5:37-39.

The USPTO issued the '675 patent over Schuster. During prosecution, the Examiner considered Schuster and specifically allowed the challenged claims over Schuster. EX-2016, ¶¶65-67.

F. Kristjansson (EX-1013)

Kristjansson discloses a device that uses distance to determine sound pressure where the user is located. EX-2016, ¶¶68-69 (citing EX-1013, Abstract).

VI. THE CHALLENGED CLAIMS OF THE '675 PATENT ARE VALID

A. Deficiencies Present in All Grounds

1. Equating Measurement to the Object Being Measured

All grounds in the Petition fail because they equate the concept of an object being classified (or determined) with the method of making that classification (or determination). Dr. Stern repeatedly used the term “measuring,” because that is all the prior art does. It simply discloses manners of measuring variables, like distance:

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

EX-2017, 95:7-17 (emphasis added). Dr. Stern testified that in the context of the prior art, “we never talk about classifying a voice. We talk about classifying based on distance.” *Id.*, 119:13-19. Even if the word “classification” could be stretched to

the ends Petitioners seek, the Petition still fails because as Dr. Stern admitted, the primary references at best only classify *distances* and not voices. EX-2017, 48:3-7, 49:7-15, 117:1-21, 118:2-13, 119:6-22.

The '675 patent discloses the concept of “volume” as distinct from “voice.” The patent states that “[t]he input unit can include a function other than the voice detecting unit 10 and, for example, a switch to adjust *volume* of the voice V2 by operation performed by the user H, and the like can be provided.” EX-1001, 3:33-36 (emphasis added); *see also id.*, 10:9-46. It also distinguishes between “voice” and *how* it is classified:

Moreover, the voice classifying unit 42 can perform the classification to the first voice V1A and the second voice V1B by using either one of a *volume* of the voice V1, a speaking speed of the user in the voice V1, and a volume ratio between a speech of a person and a wind noise as a feature value.

Id., 10:27-32 (emphasis added).

None of the prior art discloses classifying a voice separate from measuring a characteristic of the voice.

In terms of obviousness, the Petition's proffered motivations to combine do not address the *reason why* a POSITA would have modified the prior art as claimed. At most the Petition proposes that adding additional input processing to Shin or Shimomura would improve the device. But none of the motivations relate

to the full context of the claim. None relate to classifying a voice separate and distinct from just simply responding to changes in measurements (whether due to the user or change in position). *See, e.g.*, EX-2017, 29:2-30:6, 34:11-35:4, 37:10-22, 55:8-16 (e.g., “results of measurements” of Iwase and all statements of “measurement”).

Dr. Stern explained that his view is based upon the mere fact that a device can respond to changes in a whisper even if nothing else changes:

A. What I just stated was that you don't have to identify the you don't have to detect the presence of whispering in order to process whispered speech you just do the same thing you do for any other speech and again, that is what is done normally.

Id., 91:13-18.

Further, as discussed above, much of the prior art identifies that such measurements or processing can be completely independent of one another (e.g., Shin (disclosing Table 3 as distinct from Tables 1-2 or 4-9), Shimomura (disclosing Figure 7 as distinct from Figures 9-12), and the distinct modifications discussed above in connection with Aoyama and Iwase). The Petition never bridges the gap in the context of classifying an input separate and apart from adjusting an output. There is no basis upon which the Board could conclude that a POSITA would have combined input measurements to accomplish the claimed voice classification. Dr. Stern testified that he did not know whether various measurements in even a single

reference (Shin) could have worked together. *See* EX-2017, 56:15-57:5 (“I don’t know.”) In his view, that is not relevant. His conclusion is wrong. EX-2016, ¶¶71-75. To support obviousness, a POSITA must have been motivated to modify prior art within the full context of the claim.¹

Further, Shin and Shimomura disclose specific analyses of voice in simplistic fashion. For example, each relies on a single-variable table. *See, e.g.*, Pet., 13 (citing EX-1003, Table 3), 50 (citing EX-1004, Fig. 7). The Petition provides no basis explaining why a POSITA would modify the approach disclosed in connection with a process of voice classification. EX-2016, ¶76. Every stated motivation to combine presupposes that the primary reference (Shin or Shimomura) conducts a classification. *Id.* Thus, as a matter of law, the Petition

¹ *Jones v. Hardy*, 727 F.2d 1524, 1530 (Fed. Cir. 1984); *Ruiz v. A.B. Chance Co.*, 357 F.3d 1270, 1275 (Fed. Cir. 2004) (“The ‘as a whole’ instruction in title 35 prevents evaluation of the invention part by part. ... This form of hindsight reasoning, using the invention as a roadmap to find its prior art components, would discount the value of combining various existing features or principles in a new way to achieve a new result—often the very definition of invention.”); *Princeton Biochemicals, Inc. v. Coulter, Inc.*, 411 F.3d 1332, 1337 (Fed. Cir. 2005); *Allergan, Inc. v. Apotex, Inc.*, 754 F.3d 952, 966 (Fed. Cir. 2014).

provides no reasons why a POSITA would have relied on the secondary references to provide the recited classification. It certainly fails to state any motivation a POSITA would have had to modify or change a *classification* process disclosed in the primary reference. For example, even if the Board were to conclude that Shin “classifies” voice based upon measuring distance and controlling output, all obviousness grounds still fail because there is no indication that a POSITA would have been motivated to change how that process was done. As just one example, the Petition does not explain *why* a POSITA would have used the feature in connection with voice classification as opposed to adding disclosures from the secondary references as a *separate* feature. That is, an output adjustment separate and apart from any purported classification (e.g., an adjustment independent of any change in voice). EX-2016, ¶77.

2. Failure to Address Obviousness Under Prevailing Precedent

The Institution Decision relied upon the same type of conclusory obviousness rationales that the Federal Circuit has rejected.

The Board relied on two examples:

the proposed combination is merely the addition of one known element (i.e., “Aoyama’s natural language processing and information retrieval”) to another known element (i.e., “Shin’s voice	the proposed combination also represents the addition of one known element (i.e., “Aoyama’s word replacement” feature) to another known element (i.e., “Shin’s distance-
---	--

processing and response device”) to obtain predictable results (i.e., “identifying a user’s intent and providing information responsive to such intent”).	based content abbreviation/simplification”) to obtain predictable results (i.e., “using word replacement to simpl[if]y content for distant users”).
---	---

Dec., 26-27 (excerpted).

Both are incomplete. The combination must at least be made “according to known methods.” MPEP § 2143. And there is no explanation of what the known method would be. That is one example. *KSR* requires flexibility, but it does not allow for pro forma rationales reduced to simply identifying elements known in the art. *Unigene Labs., Inc. v. Apotex, Inc.*, 655 F.3d 1352, 1360 (Fed. Cir. 2011) (“Obviousness requires more than a mere showing that the prior art includes separate references covering each separate limitation in a claim under examination.”). Dr. Stern stated that he does not even know if the various modifications could work. EX-2017, 56:15-57:5 (“I don’t know.”). Further, neither rationale above addresses the full context of the claims because neither addresses classification of voices separate from changing outputs. EX-2016, ¶¶78-79.

The deficiencies do not end there. The first rationale does not address any actual difference between the claims and the prior art (under *Graham* factor two). It simply mentions garden variety natural language processing (as the difference)

and then skips to a final result (identifying and responding to user intent). The Petition's second purported rationale is a bit closer, but it still addresses "word replacement" outside the context of the claims. And it fails to explain whether the proposed combination would have any benefit, such as whether it would or could "simplify" content. In short, those statements, like others in the Petition, are too generalized to apply to either the claims at issue or the specific differences between them and the primary references (Shin and Shimomura). EX-2016, ¶80.

Additionally, neither the above statements, nor any proposed motivations provided in the Petition, identify *why* a POSITA would have been motivated to combine or modify the prior art. *Id.*, ¶81. The statements merely state that the motivation would be *to* obtain predictable results. Both statements invoke the combining-a-known-element-rationale, but they ignore identifying a reason *why* a POSITA would have been so motivated. The Supreme Court is unequivocal on this point: "A patent composed of several elements is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S., 398, 418 (2007). Surely, "common sense directs one to look with care at a patent application that claims as innovation the combination of two known devices according to their established functions" *Id.* But that is not enough. Something more is required. *Microsoft Corp. v. Enfish, LLC*, 662 F. App'x 981, 990 (Fed. Cir. 2016) ("that the references were directed to

the same art or *same techniques*” (emphasis added) was insufficient motivation to combine). *KSR* did not remove the requirement of a reason why. It simply stated that the manner of demonstrating a reason cannot be constrained to “rigid and mandatory formulas.” *KSR*, 550 U.S. at 419. Whatever the manner of demonstration, the end result must be the same: a reason why. The Board’s Decision, and the Petition, only asserts that (1) the elements would have been known and (2) that the combination would have been predictable. That is legally insufficient.

3. Reliance on Hindsight

Dr. Stern admitted to relying on “20/20 hindsight” in modifying the references. EX-2017, 59:9-60:5, 67:7-15. That fact alone undermines the opinions in his declarations and the grounds in the Petition that rely on those opinions.²

It also demonstrates why the Board’s preliminary decision (on institution) cannot support an ultimate ruling of invalidity. The Board stated that Amazon need

² See, e.g., *In re Omeprazole Patent Litigation*, 536 F.3d 1361, 1380 (Fed. Cir. 2008) (“the district court reasonably concluded that a person of ordinary skill in the art would not have seen any need to apply to [the primary reference] the teachings of the [secondary] references disclosing subcoatings” where the primary reference did not suggest any problem with the coating it had.).

not “identify a problem a [POSITA] would have sought to solve when combining the references” Dec., 27 (citing POPR, 31-32). As a preliminary matter, that statement cited only a portion of Patent Owner’s statements—even though Patent Owner delineated them as (1) and (2). POPR, 31-32. And it did not consider the full extent of the identified deficiencies. *Id.*, 33-40 (positions not yet considered by the Board).

Petitioners cannot meet their burden to provide *some* reason for the proposed combinations. And yet, that is the only way to avoid the clear hindsight bias that Petitioners and their expert used to build the Petition. For example, a combination may be the result of improper hindsight where the secondary reference provides an advantage not relevant to the primary reference. *See, e.g., In re Omeprazole Patent Litig.*, 536 F.3d at 1380. As another example, a combination may be the result of improper hindsight where the secondary reference provides only redundant advantages. *Kinetic Concepts, Inc. v. Smith and Nephew, Inc.*, 688 F. 3d. 1342, 1369 (Fed. Cir. 2012) (where two “references independently accomplish[ed] similar functions, namely, draining fluids” and the “device[s] in each reference independently operate[d] effectively, a person having ordinary skill in the art, who was merely seeking to create a better device to drain fluids from a wound, would have no reason to combine the features of both devices into a single device.”)

There is no basis in the Petition upon which the Board could conclude that a POSITA would have understood the combination of references as providing any benefit.

4. Dr. Stern's Testimony Is Unreliable

Dr. Stern's declaration is unreliable. It is not rooted in the thorough and searching analysis, with specific rational underpinnings, that the Federal Circuit requires. His positions are thus malleable, and he defaults to inferring certain disclosures in the prior art without any basis.

For example, Dr. Stern testified at his deposition that Iwase could be relied upon to disclose the distance calculation aspect of the '675 patent claims, even though that possibility was not described in his declaration. EX-2017, 103:1-16; *see also id.*, 102:9-10 ("it doesn't have to be that way"). In fact, Dr. Stern stated that any attempt to explain why he relied on certain combinations instead of others would be "foolish." *Id.* That may be accidental colloquial hyperbole. But it still demonstrates that he never proceeded through the required analysis. Otherwise the "articulated reasoning with some rational underpinnings"³ indicating an inquiry that is "thorough and searching"⁴ would be evident in the Declaration and Petition.

³ *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006).

⁴ *Palo Alto*, 122 F.4th at 1384.

Dr. Stern routinely mischaracterized its disclosure. For example, he incorrectly interpreted the '675 patent to disclose that language spoken to the device is "always in the form of a question." EX-2017, 26:21-27:5.

Dr. Stern admitted that he had not considered the full disclosure of Shin in his analysis of the reference. *Id.*, 55:8-57:5.

Dr. Stern admitted that he injected his own inferences into the analysis to supplement the disclosure of Shin, beyond the actual disclosure of the reference. *Id.*, 57:18-59:3 ("I added [the annotations] ... to demonstrate that distance measurements like these *could be* used to classify voices as in the '675 patent" (emphasis added)). Tellingly, Dr. Stern testified that Shin did not actually disclose classifying voices, as required by the '675 patent. *Id.*, 48:3-7, 49:7-15.

Dr. Stern incorrectly relied on Shin's disclosure of classifying *content* "dichotomously" to support his conclusion that Shin disclosed classifying *voices* "dichotomously." EX-1002, ¶61 (citing EX-1003, ¶[0051]). He admitted that paragraph [0051] of Shin only discloses classifying *output content* dichotomously. EX-2017, 59:4-60:12, 62:3-63:18, 65:11-22.

Dr. Stern incorrectly linked various distance ranges to different voice categories when analyzing Shin. *Id.*, 72:6-75:14 ("Shin doesn't infer, I infer."). The association of *voice* with a *distance* is not disclosed in Shin, only the association of a user with distance.

Dr. Stern admitted that he injected his own inferences to supplement the disclosure of Shimomura. *Id.*, 119:6-22, 120:3-13 (“I put [the annotations] [] in ... [as an] attempt to establish a correspondence between the classifications performed by the Shimomura device and the ’675 invention”). Tellingly, Dr. Stern testified that Shimomura does not actually disclose classifying voices, as required by the ’675 patent. *Id.*, 117:1-21, 119:6-22.

Dr. Stern admitted that Iwase does not actually disclose calculating a distance. *Id.*, 96:8-9 (“there wasn’t necessarily a distance measurement here”).

And Dr. Stern admitted his bias in favor of the Petitioners. He has a “long-time interaction with Amazon,” so much that he provides his services at a discounted rate. *Id.*, 10:4-10.

B. GROUND 1A: Shin in View of Aoyama Does Not Render the Challenged Claims Obvious

1. Claims 1-7: Shin Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice

As discussed in Section IV.A above, independent claims 1, 6, and 7 require classifying *voices* (i.e., “a first *voice* or a second *voice*”) based on calculated distance. EX-1001, claims 1, 6-7. For Ground 1A, the Petition relies solely on Shin in connection with the voice classification limitation. Pet., 12-15. However, while Shin discloses determining output schemes based on distance ranges, it does not disclose that any such range is sufficient for the purpose of classifying voices. Dr.

Stern confirmed as much. He testified, in connection with Shin's output, that "the decision about which form to present it in is not based on the incoming speech. It is based on the distance between the user of the device." EX-2017, 65:10-22. The Petition acknowledges that Shin "adjusts the amount of information provided *based on the user-to-device distance.*" Pet., 22 (emphasis added). 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 recites language, in conclusory fashion (the portion in italics): "Shin's *processor analyzes the voice to classify it as either a first or second voice* based on the user-to-device distance." Pet., 12 ((citing EX-1002, ¶60) (emphasis added) (citing only expert declaration)).

The Petition also contradicts Shin. The only express disclosure of "voice" the Petition identifies in Shin relates to processing voice in connection with "generat[ing] . . . content." Pet., 12. That is, garden variety natural language processing to determine user intent. EX-2016, ¶¶82-83. None of the sentences that follow in that section point to any express disclosure of "voice" or "classification" in Shin. The only exception occurs in paragraph 51 of Shin. But that paragraph only mentions "classification" of "content." EX-1003, ¶[0051]. And Shin mentions that classification of content solely in the context of output, and not voice input. 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 (Dr. Stern) or the '675 patent. Neither the Petition nor Dr. Stern provides any reasoned analysis supporting the conclusion that Shin classifies voices.

The Petition then turns to Table 3 of Shin. Specifically, it includes an annotated version of the table, where Petitioners' expert added the "First Voice" and "Second Voice" labels. Pet., 13 (citing EX-1002, ¶61). But annotations aside, Table 3 merely 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)

EX-1003, Fig. 3.

That determination of distance in Table 3 cannot expressly disclose a classification, because it is a measurement of a continuous variable, not a discrete classification. EX-2016, ¶¶84-86. Nor can it expressly disclose an analysis of a "voice," which is not at all synonymous with "distance." *Id.*, ¶86. One example is sufficient to demonstrate the failed reasoning here. *Id.*, ¶87. 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.” 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.” EX-1003, ¶[0053]. Thus, a POSITA would never have understood Shin’s disclosure of distance to be synonymous with “voice.” EX-2016, ¶88. Nor would a POSITA have understood Shin’s determination of a distance to be synonymous with classification. *Id.*

Moreover, Shin directly correlates content output to distance between the user and the device. It does not classify the voice input in any way, let alone classify the voice input as a first voice or a second voice. Petitioners tacitly acknowledge that Shin’s classes/groupings are based on distance, not voices. Pet., 13 (“*Any distance range beyond 1 meter* would satisfy the claimed ‘first voice.’

Alternatively, all distances over 1 meter could collectively be a first voice (e.g., a voice greater than 1 meter away).” (citations omitted)). In contrast, as discussed in Section IV.A, 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, 10:4-8; EX-2016, ¶89.

Crucially, Shin’s Table 3 does not even disclose voice classification under Dr. Stern’s own understanding. Specifically, he testified in a related IPR that whether a classification constitutes a *voice* classification depends on the input and output of the classification. *See* EX-2020, 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. *See also* EX-2017, 49:7-15; EX-2016, ¶90.

Dr. Stern 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.

See EX-2020, 72:12-73:8 (emphasis altered); EX-2016, ¶91.

Accordingly, the Petition offers 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). *Id.*, ¶92.

Thus, Shin does not disclose or suggest "classify the voice into either a first voice or a second voice based on the calculated distance" as claimed and described by the specification. *Id.*, ¶93.

2. Claims 1-7: Shin in View of Aoyama Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice

Independent claims 1, 6, and 7 recite: "generat[e/ing], when the voice is determined to be the first voice, a first output sentence in which at least one word

selected among words included in the content information of the notification information is replaced with another word.” EX-1001, claims 1, 6-7. The Petition alleges that Shin in view of Aoyama discloses this limitation. Not so.

The Petition alleges that Shin discloses user-to-device distance ranges at which the output content is abbreviated. Pet., 22-23. The Petition further alleges that “[a] POSITA would have readily understood Shin’s disclosure of abbreviating content to describe it more simply or provide the ‘gist’ includes replacing words.” *Id.* (citing EX-1002, ¶92). The Petition provides no reasoned basis for that assertion; it is purely conclusory. EX-2016, ¶¶94-95. Shin provides no indication that the word “abbreviation” means or includes “replacing words.” *Id.*, ¶95.

In the context of Shin, abbreviating content involves excluding certain portions of output content. Shin discloses, by way of example, that an original content may include “[t]he weather in Jul. 1, 2015 is (1) rainy after cloudy (2). The highest temperature is 28° C., and the lowest temperature is 18° C. (3), the rainfall is 10 mm (4). Prepare your umbrella when you go out. (5).” EX-1003, ¶[0088]. Shin further discloses that:

“the content of a first abbreviation level may include information of (1), (2), (3), and (4) of the original content while excluding (5), and the content of a second abbreviation level may include information of (1), (2) and (4) while excluding (3) and (5). Furthermore, the content of a third abbreviation level may only

include information of (1) and (2), and the content of a fourth abbreviation level (abstract content) may only include information of (2).”

Id. Thus, Shin discloses abbreviating content by excluding certain information included in the original content. *Id.*; EX-2016, ¶¶96-97. Shin nowhere discloses abbreviating content by replacing words. *Id.*, ¶97. For that reason, Petitioners introduce an alternate theory. Pet., 23 (“Even if Shin did not expressly disclose replacing words to generate abbreviated/simplified content”)

The Petition alleges that replacing words would have been obvious over Aoyama. Pet., 23-25. Specifically, the Petition alleges that “Aoyama’s device tailors output by replacing words based on the user or situation.” Pet., 23 (citing EX-1010, ¶¶[0120]-[0124], [0138]-[0140], [0177], [0218]). However, Aoyama teaches inputting words into a response based on the “speech style of the user.” EX-1010, ¶[0121]; EX-2016, ¶98. It does not make any adjustment based on whether the user’s voice is determined to be a “first voice” or “second voice,” as claimed. EX-2016, ¶98. Specifically, Aoyama discloses that “speech style detection unit 115” extracts a user’s “unique expressions from text information obtained by converting the sound input on the basis of the result of the natural language processing on the text information.” EX-1010, ¶¶[0086]-[0089]. For example, a user may refer to the name “Taro Yamada” by different expressions

such as “Mr. Yamada,” “Taro,” or “Taro Yamada.” *Id.* Aoyama’s speech style detection unit 115 extracts a user’s speech style, including the user’s unique expressions for personal names, date, and time. *Id.* Aoyama’s response parameter generation unit 16 generates a response parameter based on “speech style of the user,” and notifies response generation unit 17 of words to be included in the generated response. *Id.*, ¶¶[0121]-[0124]. In this way, the “response generation unit 17 can generate a response in which the person with the name ‘Taro Yamada’ is called ‘Mr. Yamada’ in accordance with the speech style of the speaker.” *Id.* Thus, Aoyama chooses a word (e.g., “Mr. Yamada”) based on a user’s unique expression. *Id.* Dr. Stern incorrectly interpreted the determination of “user states” to be a “classification” as used and described in the ’675 patent. EX-2017, 21:7-13. The Petition does not allege, and Aoyama does not disclose, the claimed voice classification. EX-2016, ¶98. Nor does it provide any basis for a word replacement based on a voice classification. *Id.*

Aoyama’s word choices are based on the “speech style of the user,” whereas Aoyama’s change in “amount of information output” (similar to Shin’s content abbreviation) is based on the user’s “degree of composure.” EX-1010, ¶¶[0121]-[0124], [0145]-[0147]; EX-2016, ¶99. Aoyama teaches that those features involve separate and distinct processes and metrics. EX-2016, ¶99. Accordingly, a POSITA seeking to incorporate Aoyama’s word input feature into Shin’s device

would not have modified Shin to replace words in Shin's output content based on user-to-device distance. *Id.* Aoyama teaches that words are *only* to be chosen in response to "speaking style." EX-1010, ¶[0049]; EX-2016, ¶99. That "speaking style" is tied to the specific words a user tends to use. EX-2016, ¶99. Aoyama, for example, distinguishes between users who tell time based on a 12-hour clock (e.g., "3:00") versus those who speak in military time (e.g., "15:00"). *Id.* A POSITA would have understood Aoyama as teaching away from replacing words based on distance. *Id.* Indeed, doing so would have defeated the purpose of Aoyama. *Id.* It would have led to a response with which "the user may feel uncomfortable." EX-1010, ¶[0049]; EX-2016, ¶99. For example, combining Shin and Aoyama's disclosures would have the device change from responding in standard time to military time just because a user moved closer to or further away from the device. EX-2016, ¶99. And there is no indication a POSITA would have known that it could have worked, let alone have been motivated to do so. *See* EX-2017, 56:15-57:5; EX-2016, ¶99.

Even *if* a POSITA would have been motivated to incorporate Aoyama's word replacement feature into Shin's device, such a skilled person nonetheless would not have been motivated to combine Shin and Aoyama in the manner claimed. EX-2016, ¶100. Like Shin, Aoyama discloses changing the "amount of information output as the response." EX-1010, ¶¶[0145]-[0147]; EX-2016, ¶100.

For example, Aoyama discloses that for a sound input “Show me how to get to Tokyo Station,” “the response generation unit 17 may output only ‘transfer information regarding how to get to Tokyo Station’ indicating the response content when the user is in a hurried state.” EX-1010, ¶¶[0145]-[0147]. In contrast, “when the user is calm, the response generation unit 17 may output ‘weather near Tokyo Station’” as additional information. *Id.* Thus, Aoyama discloses changing the “amount of information output as the response” based on a user’s “degree of composure.” *Id.*; EX-2016, ¶100. Aoyama only discloses inputting words based on the “speech style of the user,” as discussed above. EX-2016, ¶100. Thus, if anything, a POSITA would have combined the information-modifying portion of Aoyama with Shin’s information-modifying portion. *Id.* And, based on Aoyama’s teachings, that would have remained distinct from word replacement. *Id.*

3. Claim 5: Shin Does Not Disclose Determining the Voice Spoken by the User is a Whisper

Claim 5 (which depends from claim 1) recites: “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” EX-1001, claim 5.

The Petition alleges that Shin discloses this limitation because “a POSITA would have understood Shin’s disclosure to encompass *receiving and processing as wide a range of speech as possible, including whispered speech.*” Pet., 38. The

Petition further alleges that “[e]ven if Shin’s device did not explicitly support *processing whispered speech*, a POSITA would have been motivated for several reasons to configure the device to *receive and process such inputs*.” *Id.* Those positions are facially deficient.

The Petition asserts that a POSITA would have understood a whisper to be a subset of “a range of speech volumes and rates.” *Id.* There is no basis for equating a whisper to be a range of volumes and rates. EX-2016, ¶¶101-103. At best, that position conflates the object being classified with the manner of classifying it. EX-*Id.*, ¶103. Further, Petitioners’ approach would vitiate any difference in scope between claim 1 and 5. *Id.*

Dr. Stern confirmed that Amazon’s Petition is based upon ignoring the plain meaning of the claim:

Q Okay. So your position is that claim 5 does not require the determination that a voice is a whisper; is that correct?

A That’s correct.

EX-2017, 93:15-18.

A POSITA would have understood that the recited output-content control device could not be “configured to” determine a predetermined voice “when the voice spoken by the user is a whisper,” if it is not capable of identifying a whisper. EX-2016, ¶104. It would not be configured as claimed. *Id.* Dr. Stern’s application

of Shin to disclose claim 5 is based on that faulty claim interpretation, and an overly broad interpretation of Shin. EX-2017, 90:2-91:5 (“Shin’s device as described is configured to work over a wide range of conditions, volumes, et cetera, et cetera, which means that it handles whispered speech as well.”).

The Petition acknowledges the lack of any whisper detection by citing a reference (Iyer) that does not form basis of the combination in Ground 1A—or any other prior art combination in the Petition. The Board recognized in the Institution Decision that doing so is improper. Dec., 22 n.5 (citing, among others, *Randall Mfg. v. Rea*, 733 F.3d 1355, 1362–63 (Fed. Cir. 2013) (emphasizing that additional prior art references or evidence are not for the purpose of changing the prior art combination that forms the basis of the asserted ground”)). A reference can be cited to establish background of knowledge but not to modify the references in the combination.⁵ Thus, Ground 1A fails for that reason alone—the Petition does exactly what *Randall* prohibits. It does not provide Iyer as context. It provides Iyer to supply the missing limitation of determining a “whisper.”

⁵ See USPTO Memorandum re Rule 104(b)(4), dated July 21, 2025

(https://www.uspto.gov/sites/default/files/documents/aapa_memo_final_signed.pdf?utm_campaign=subscriptioncenter&utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term=).

Further, the Petition does not identify any express or inherent disclosure in Shin that allegedly meets the elements of claim 5. EX-2016, ¶¶105-106. Instead, it states that Shin “encompass[es] *receiving and processing ... whispered speech*,” or that it would have been obvious to configure Shin’s device “to *receive and process [whispered speech]*.” Pet., 38. But claim 5 requires more than the capacity to receive and process whispered speech. EX-2016, ¶106. It requires a *determination* that “the voice spoken by the user is a whisper.” *Id.* That determination cannot be made if a whisper cannot be identified. The specification makes that clear:

The voice classifying unit 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and determines whether the voice is a predetermined voice. In the present embodiment, the voice classifying unit 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and classifies the voice V1 to either one of a first voice V1A and a second voice V1B. *The first voice V1A corresponds to a predetermined voice.* In the present embodiment, the voice classifying unit 42 classifies the voice V1 to the first voice V1A *when the voice V1 is determined to be a whisper as a result of analysis* (determines as the predetermined voice). The voice classifying unit 42 classifies the voice V1 to the second voice V1B when the voice V1 is determined not to be a whisper.

EX-1001, 9:58-10:8 (emphasis added); EX-2016, ¶106.

Ground 1A thus fails to provide any tenable ground of invalidity.

4. Claims 1-7: Shin Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice

Claims 1, 6, and 7 recite generating output sentences based on whether the “voice is determined to be” the “first voice” or the “second voice.” EX-1001; *see also* EX-2016, ¶107.

For reasons discussed in Sections VI.B.1 above, Shin does not disclose or suggest classifying the voice as the first and second voice as claimed. *Id.*, ¶108. For the same reasons, Shin does not disclose or suggest generating a first and second output sentence when the voice has been classified as the first and second voice. *Id.*

Dr. Stern relied on the “original content” of Shin to disclose the claimed requirement in the '675 patent of a “second output sentence which includes all of the intention information and the content information.” EX-1002, ¶106 (citing EX-1003, ¶¶[0051], [0087]-[0088]). However, when faced at his deposition with the reality that Shin actually did *not* directly disclose that, Dr. Stern attempted to supplement the actual disclosure of Shin—and his declaration—with his own assumptions in order to meet the claim element. EX-2017, 75:16-82:15 (e.g., “the intention information is already *implicitly* in there”).

Aoyama does not cure the deficiencies of Shin. EX-2016, ¶108. Therefore Ground 1A fails to invalidate any claim.

5. Claims 1-7: Shin Does Not Disclose the Claimed Components Arranged as in the Claims

For Ground 1A, the Petition focuses on the claimed functions without identifying what *component* allegedly accomplishes those functions. EX-2016, ¶109. Ground 1A does not address the details or arrangement of the components/steps of any claim of the '675 patent. Amazon tacitly admits that deficiency in, e.g., Ground 5. Petitioners introduce Kristjansson in Ground 5 to assert disclosure of certain recited components that perform certain functions. Pet. 67 (addressing “intention analyzing unit” and “notification-information acquiring unit”). But even then, Amazon's attempt to cure the deficiencies of Ground 1A is cursory. The Petition spends most of its pages here repeating incomplete obviousness rationales. Pet., 67-69; EX-2016, ¶109.

In short, Petitioners admit, via Ground 5, the necessity to identify the detailed arrangement of components/steps recited in the other claims. But they failed to do so in Ground 1A. EX-2016, ¶110.

6. Claims 2-5

Claims 2-5 depend from claim 1. EX-1001, claims 2-5. Therefore, for at least the reasons discussed above for claim 1, Ground 1A fails to invalidate dependent claims 2-5. EX-2016, ¶111.

C. GROUND 1B: Shin in View of Iwase Does Not Render the Challenged Claims Obvious

1. Claims 1-7: Shin Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice

Shin does not disclose the claimed voice classification for the same reasons discussed in Section VI.B.1. Iwase does not cure the deficiencies of Shin on this claim limitation, and the Petition does not rely on Iwase to disclose it. EX-2016, ¶112. Therefore, Ground 1B fails to invalidate any claim.

2. Claims 1-7: Shin in View of Iwase Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice

Independent claims 1, 6, and 7 recite: “generat[e/ing], when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word.” EX-1001, claims 1, 6-7. For Ground 1B, the Petition alleges that Shin in view of Iwase would have rendered this limitation obvious. For the same reasons discussed in Section VI.B.2, Shin does not disclose it. EX-2016, ¶113.

The Petition alleges that Iwase discloses or suggests this limitation because Iwase's device adjusts notification information to users “based on user attributes, such as age or word preferences,” which “may involve word replacement, such as replacing ‘I recommend’ in the phrase ‘I recommend bringing an umbrella’ with

‘You should,’ yielding: ‘You should take an umbrella.’” Pet., 41-42 (citing EX-1017, 1:54-2:18, 4:17-29, 6:32-35, 12:26-42, 15:6-40); EX-2016, ¶114. But Iwase does not disclose replacing words based on voice classification. EX-2016, ¶114.

As a threshold matter, Iwase does not expressly disclose replacing words in original output content. The word “replace” does not appear in Iwase. Neither the Petition nor Petitioners’ expert explains where in Iwase a POSITA would have understood there to have been a replacement. Nor does it provide any basis for an inherency position to support such a conclusion. EX-2016, ¶115.

Further, Iwase does not disclose word replacement based on voice classification as required by the challenged claims. It discloses that “speaking control unit 360 shapes the extracted subject on the basis of *attribute information of a user*.” EX-1017, 15:4-10. That attribute information includes “age, distinction of sex, language, and hometown.” *Id.*, 12:27-34. Iwase does not disclose any operation based on voice classification. EX-2016, ¶116. It discloses adjusting outputs based on an “attention acquisition difficulty level,” which is calculated based on “noise level and a situation of the user,” as well as “attribute information of a user.” EX-1017, 10:35-67, 12:27-42, 15:41-45, Fig. 7; EX-2016, ¶116.

“Attributed information of a user” in Iwase includes, for example, age, distinction of sex, language, and hometown.” EX-1017, 12:26-31. None of that involves any voice analysis, let alone voice classification. EX-2016, ¶116.

Accordingly, a POSITA would have no reason to combine Shin and Iwase in the manner claimed. EX-2016, ¶117. Iwase is not concerned at all with distance. *Id.* It does not improve any process that incorporates distance. *Id.* The Petition provides no basis for asserting that a POSITA would have been motivated to make the combination. *Id.* The Petition's sole rationale is that Iwase's feature would allow a user "to grasp notification contents more effectively." Pet., 42 (citing 2:11-14, 15:24-28, 21:66-22:8). That is an empty, conclusory assertion. EX-2016, ¶117. Neither Petitioners nor their expert explain how replacing one word with another word at a certain user-to-device distance would have furthered the goal of either reference. *Id.* It reduces to purely circular reasoning: that a POSITA would have combined Shin and Iwase to arrive at the combination. Pet. 42-44; EX-2016, ¶117. Nor does the Petition identify any manner in which a POSITA would have recognized Shin to need improvement— certainly not by any method that excludes the use of distance. EX-2016, ¶117. For example, the Petition does not provide any analysis of why a POSITA would have desired to add complexity to Shin, when the abbreviated information sufficed for Shin's purposes. *Id.*; see *Palo Alto*, 122 F.4th at 1384 ("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." (cleaned up)).

Therefore, Ground 1B fails to invalidate any claim.

3. Claim 5: Shin Does Not Disclose Determining the Voice Spoken by the User is a Whisper

The combination of Shin and Iwase does not disclose or suggest claim 5 for the same reasons discussed in Section VI.B.3. Iwase does not cure the deficiencies of Shin on this claim, and the Petition does not rely on Iwase to disclose it. EX-2016, ¶118. Therefore, Ground 1B fails to invalidate any claim.

4. Claims 1-7: Shin Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice

Independent claims 1, 6, and 7 recite: “generat[e/ing], when the voice is determined to be the first voice, a first output sentence, ... and to generat[e/ing], when the voice is not determined to be the second voice, a second output sentence.” EX-1001, claims 1, 6, 7. For reasons discussed in Sections VI.B.4, Shin does not disclose this limitation. Iwase does not cure the deficiencies of Shin on this claim, and the Petition does not rely on Iwase to disclose it. Therefore, Ground 1B fails to invalidate any claim. EX-2016, ¶119.

5. Claims 1-7: Shin Does Not Disclose the Claimed Components Arranged as in the Claims

The combination of Shin and Iwase does not disclose or suggest the claimed components for the same reasons discussed in Section VI.B.5. EX-2016, ¶120.

6. Claims 2-5

Because claims 2-5 depend from claim 1, they are valid under Ground 1B for at least the reasons discussed above for claim 1. *Id.*, ¶121

D. GROUND 2A: Shimomura in View of Aoyama Does Not Render the Challenged Claims Obvious

1. Claims 1-7: Shimomura Discloses Measurement of Distance, Not Classification of Voices

As discussed in Section IV.B.1 above, all challenged claims require classes of *voices* (i.e., “a first *voice* or a second *voice*”). None of the asserted grounds and prior art therein disclose classifying a voice spoken by a user “as either a first voice [and/or] a second voice,” as claimed and described in the specification. EX-2016, ¶122.

Shimomura does not disclose any manner of classifying voices. *Id.*, ¶123. The Petition acknowledges that Shimomura discloses a “device” that “controls speech forms ... according to the *physical distance* to the conversation partner.” Pet., 47 (emphasis added) (citations omitted). The only basis the Petition provides for “classification” of a “voice” is in connection with Fig. 7. EX-2016, ¶123. The Petition does not identify any express disclosure of either aspect. *Id.* Indeed, the Petition 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 ’675 patent. *Id.* Neither the Petition nor Dr. Stern provides any

reasoned analysis supporting the conclusion that Shimomura classifies voices. *Id.*

Thus, the Petition fails to demonstrate that Shimomura discloses voice classification. *Id.*

The Petition also provides no reasoned analysis for why a POSITA would equate Shimomura's distance measurements with either the concept of a "voice" or "classification." EX-2016, ¶124. And it provides no inherency assertion.

Similarly, to the extent that Shimomura discloses any required, claimed classification, it discloses classes of distances, not voices. EX-2016, ¶125. The Petition asserts 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

Pet., 50; EX-1004, Fig. 7 (annotated by Petitioners).

Dr. Stern added the "First Voice" and "Second Voice" labels to Fig. 7. EX-2016, ¶126. Annotations aside, Figure 7 discloses one thing: a calculation of distance and nothing else. *Id.*, ¶126. 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.*

In fact, Shimomura confirms the distinction between “voice,” “distance,” and “volume.” *Id.*, ¶127. 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.

EX-1004, ¶[0004] (emphasis added).

Thus, a POSITA would never have understood Shimomura's disclosure of distance to be synonymous with “voice.” EX-2016, ¶128. Nor would a POSITA have understood Shimomura's determination of a distance to be synonymous with a classification. *Id.*

In short, like Shin, Shimomura directly correlates the output content and volume based on distance without performing the intermediate step of classifying the voice. 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 and not voices, similar to as discussed above for Shin. EX-2016, ¶129.

Therefore, Ground 2A fails to invalidate any claim.

2. Claims 1-7: Shimomura in View of Aoyama Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice

The combination of Shimomura and Aoyama does not disclose replacing words based on voice classification for similar reasons discussed in Section VI.B.2. Namely, the Petition does not assert that Shimomura discloses word replacement. EX-2016, ¶130. It relies on Aoyama for that limitation. Pet., 55. And the proposed combination of Shimomura with Aoyama fails for the same reasons. EX-2016, ¶130.

Aoyama's word choices are based on the "speech style of the user," whereas Aoyama's change in "amount of information output" (similar to Shin's content abbreviation) is based on the user's "degree of composure." EX-1010, ¶¶[0121]-[0124], [0145]-[0147]; EX-2016, ¶131. Aoyama teaches that those features involve separate and distinct processes and metrics. *Id.* Accordingly, a POSITA seeking to incorporate Aoyama's word input feature into Shimomura's device would not have modified Shimomura to replace words in Shimomura's output content based on user-to-device distance. *Id.* Aoyama teaches that words are *only* to be chosen in response to "speaking style." EX-1010, ¶[0049]; EX-2016, ¶131. That "speaking style" is tied to the specific words a user tends to use. *Id.*, ¶131. Aoyama, for example, distinguishes between users who tell time based on a 12-hour clock (e.g.,

“3:00”) versus those who speak in military time (e.g., “15:00”). *Id.* A POSITA would have understood Aoyama as teaching away from replacing words based on distance. *Id.* Indeed, doing so would have defeated the purpose of Aoyama. *Id.* It would have led to a response with which “the user may feel uncomfortable.” *Id.* For example, combining Shimomura and Aoyama would have the device change from responding in standard time to military time just because a user moved closer to or further away from the device. *Id.* And there is no indication a POSITA would have known that it could have worked, let alone have been motivated to do so. *Cf.* EX-2017, 56:15-57:5; EX-2016, ¶131.

Even *if* a POSITA would have been motivated to incorporate Aoyama's word replacement feature into Shimomura's device, such a skilled person nonetheless would not have been motivated to combine Shimomura and Aoyama in the manner claimed. EX-2016, ¶132. Like Shimomura, Aoyama discloses changing the “amount of information output as the response.” EX-1010, ¶¶[0145]-[0147]; EX-2016, ¶132. For example, Aoyama discloses that for a sound input “Show me how to get to Tokyo Station,” “the response generation unit 17 may output only ‘transfer information regarding how to get to Tokyo Station’ indicating the response content when the user is in a hurried state.” *Id.*; EX-2016, ¶132. In contrast, “when the user is calm, the response generation unit 17 may output ‘weather near Tokyo Station’” as additional information. *Id.* Thus, Aoyama

discloses changing the “amount of information output as the response” based on a user’s “degree of composure.” *Id.* Aoyama only discloses inputting words based on the “speech style of the user,” as discussed above. Thus, if anything, a POSITA would have combined the information-modifying portion of Aoyama with Shimomura’s information-modifying portion. And, based on Aoyama’s teachings, that would have remained distinct from word replacement. EX-2016, ¶132.

3. Claim 5: Shimomura Does Not Disclose Determining the Voice Spoken by the User is a Whisper

Claim 5 (depends from claim 1) recites: “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” EX-1001, claim 5.

The Petition alleges that Shimomura discloses this limitation because “Shimomura discloses that its device *receives and processes* speech input from a user,” and “whisper is a form of speech.” Pet., 58. The latter statement does not support any conclusion that processing speech includes determining whether speech input was provided by a whisper. Thus, there is no stated basis upon which one would conclude that Shimomura discloses any manner of whisper determination. *Cf.* Section VI.C.3 (lack of any basis for concluding that the manner of classifying an object equates to the object itself). Indeed, Dr. Stern admitted that he essentially read that limitation out of the claim:

Q Okay. So your position is that claim 5 does not require the determination that a voice is a whisper; is that correct?

A That's correct.

EX-2017, 93:15-18.

A POSITA would have understood that the recited output-content control device could not be “configured to” determine a predetermined voice “when the voice spoken by the user is a whisper,” if it is not capable of identifying a whisper. EX-2016, ¶¶133-135. It would not be configured as claimed. *Id.* Dr. Stern's application of Shimomura fails for the same reason as discussed in connection with Shin. EX-2017, 90:2-91:5 (“Shin's device as described is configured to work over a wide range of conditions, volumes, et cetera, et cetera, which means that it handles whispered speech as well.”).

The Petition further alleges that “Even if Shimomura did not teach receiving and processing whispered speech, a POSITA would have been motivated for several reasons to configure the device to do so.” Pet., 58. The Petition provides no support for that assertion. The Petition improperly relies on prior art not included in the proffered combination (Iyer), as discussed above in Section VI.B.3.

Claim 5 recites “determin[ing] ... the voice spoken by the user is a whisper.” EX-1001. The Petition does not identify any express or inherent disclosure in Shimomura that allegedly meets the elements of claim 5. EX-2016, ¶¶136-137.

Instead, it states that “Shimomura discloses classifying a whispered voice” because “Shimomura discloses that its device *receives and processes* speech input from a user,” and “whisper is a form of speech.” Pet., 58. But claim 5 requires more than the capacity to receive and process whispered speech. *Id.*, ¶137. It requires a determination that “the voice spoken by the user is a whisper.” *Id.* That determination cannot be made if a whisper cannot be identified. The specification makes that clear:

The voice classifying unit 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and determines whether the voice is a predetermined voice. In the present embodiment, the voice classifying unit 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and classifies the voice V1 to either one of a first voice V1A and a second voice V1B. *The first voice V1A corresponds to a predetermined voice.* In the present embodiment, the voice classifying unit 42 classifies the voice V1 to the first voice V1A *when the voice V1 is determined to be a whisper as a result of analysis* (determines as the predetermined voice). The voice classifying unit 42 classifies the voice V1 to the second voice V1B when the voice V1 is determined not to be a whisper.

EX-1001, 9:58-10:8 (emphasis added).

Again, Dr. Stern confirmed that Amazon's Petition is based upon ignoring the plain meaning of the claim. EX-2017, 93:15-18 (confirming that claim 5 does not require the determination that a voice is a whisper). EX-2016, ¶138.

Ground 2A thus fails to provide any tenable ground of invalidity.

4. Claims 1-7: Shimomura Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice

Claims 1, 6, and 7 recite generating output sentences based on whether the “voice is determined to be” the “first voice” or the “second voice.” EX-1001; *see also id.*, 20.

For reasons discussed in Section VI.D.1 above, Shimomura does not disclose or suggest classifying the voice as the first and second voice as claimed. For the same reasons, Shimomura does not disclose or suggest generating a first and second output sentence when the voice has been classified as the first and second voice. EX-2016, ¶¶139-140.

Dr. Stern also incorrectly relied on the “original content” of Shimomura to disclose the claimed requirement in the '675 patent of a “second output sentence which includes all of the intention information and the content information.” EX-1002, ¶193 (citing EX-1004, ¶¶[0057]-[0058], [0062], [0065], [0085], [0092]); EX-2017, 124:1-125:19. But there is no disclosure in Shimomura that the “original content” (which Dr. Stern mapped to the “second output sentence” claimed in the

'675 patent (EX-2017, 125:2-16, 126:4-9)) includes *both* all of the intention information and the content information. EX-2016, ¶141.

5. Claims 1-7: Shimomura Does Not Disclose the Claimed Components Arranged as in the Claims

For Ground 2A, the Petition focuses on the claimed functions without identifying what *component* allegedly accomplishes those functions. EX-2016, ¶142. Ground 2A does not address the details or arrangement of the components/steps of any claim of the '675 patent. *Id.* Petitioners tacitly admit that deficiency in Ground 5. Petitioners introduce Kristjansson in Ground 5 to assert disclosure of certain recited components that perform certain functions. Pet., 67 (addressing “intention analyzing unit” and “notification-information acquiring unit”). But even then, Amazon's attempt to cure the deficiencies of Ground 2A is cursory. The Petition spends most of its pages repeating incomplete obviousness rationales. Pet., 67-69; EX-2016, ¶142.

In short, Amazon admits, via Ground 5, of the necessity to identify the detailed arrangement of components/steps recited in the other claims. But it failed to do so in Ground 2A.

6. Claims 2-5

Because claims 2-5 depend from claim 1, they are valid under Ground 2A for at least the reasons discussed above for claim 1. EX-2016, ¶143.

E. GROUND 2B: Shimomura in view of Iwase Does Not Render the Challenged Claims Obvious

1. Claims 1-7: Shimomura Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice

Shimomura does not disclose the claimed voice classification for the same reasons discussed in Section VI.D.1. EX-2016, ¶144.

2. Claims 1-7: Shimomura in View of Iwase Does Not Disclose Replacing Words in a First Output Sentence When the Voice Is Determined to be the First Voice

The combination of Shimomura and Iwase does not disclose replacing words based on voice classification for similar reasons discussed in Section VI.D.2.

Namely, the Petition does not assert that Shimomura discloses word replacement.

EX-2016, ¶145. It relies on Iwase for that limitation. Pet., 60-62. But the

combination of Shimomura with Iwase still fails to disclose the claim limitation.

EX-2016, ¶145.

As a threshold matter, Iwase does not expressly disclose replacing words in original output content. The word “replace” does not appear in Iwase. Neither the Petition nor Petitioners' expert explains where in Iwase a POSITA would have understood there to have been a replacement. Nor does it provide any basis for inherency. *Id.*, ¶146.

Further, Iwase does not disclose word replacement based on voice classification as required by the challenged claims. It discloses that “speaking

control unit 360 shapes the extracted subject on the basis of *attribute information of a user*.” EX-1017, 15:4-10. That attribute information includes “age, distinction of sex, language, and hometown.” *Id.*, 12:27-34. Iwase does not disclose any operation based on voice classification. It discloses adjusting outputs based on an “attention acquisition difficulty level,” which is calculated based on “noise level and a situation of the user,” as well as “attribute information of a user.” *Id.*, 10:35-67, 12:27-42, 15:41-45, Fig. 7; EX-2016, ¶147. “Attributed information of a user” in Iwase includes, for example, age, distinction of sex, language, and hometown.” EX-1017, 12:26-31. None of that involves any voice analysis, let alone voice classification. EX-2016, ¶147.

Accordingly, a POSITA would have no reason to combine Shimomura and Iwase in the manner claimed. Iwase is not concerned at all with distance. It does not improve any process that incorporates distance. *Id.*, ¶148. The Petition provides no basis for asserting that a POSITA would have been motivated to make the combination. *Id.* The Petition's sole rationale is that Iwase's feature would allow a user “to grasp notification contents more effectively.” Pet., 42 (citing 2:11-14, 15:24-28, 21:66-22:8). That is an empty, conclusory assertion. EX-2016, ¶148. Neither Petitioners nor their expert explains how replacing one word with another word at a certain user-to-device distance would have furthered the goal of either reference. *Id.* It reduces to purely circular reasoning: that a POSITA would have

combined Shimomura and Iwase to arrive at the combination. Pet., 42-44; EX-2016, ¶148. Nor does the Petition identify any manner in which a POSITA would have recognized Shimomura to need improvement—certainly not by any method that excludes the use of distance. EX-2016, ¶148. For example, the Petition does not provide any analysis of why a POSITA would have desired to add complexity to Shimomura, when the adjusted speech intonation sufficed for Shimomura's purposes. *See Palo Alto*, 122 F.4th at 1384; EX-2016, ¶148.

3. Claim 5: Shimomura Does Not Disclose Determining the Voice Spoken by the User is a Whisper

The combination of Shimomura and Iwase does not disclose or suggest claim 5 for the same reasons discussed in Section VI.D.3. EX-2016, ¶149.

4. Claims 1-7: Shimomura Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice

Independent claims 1, 6, and 7 recite: “generat[e/ing], when the voice is determined to be the first voice, a first output sentence, ... and to generat[e/ing], when the voice is not determined to be the second voice, a second output sentence.” EX-1001, claims 1, 6-7. For reasons discussed in Sections VI.D.4, The Shimomura and Iwase combination does not disclose this limitation. Iwase does not cure the deficiencies of Shimomura on this claim, and the Petition does not rely

on Iwase to disclose it. Therefore, Ground 2B fails to invalidate any claim. EX-2016, ¶¶150-151.

5. Claims 1-7: Shimomura Does Not Disclose the Claimed Components Arranged as in the Claims

The combination of Shimomura and Iwase does not disclose or suggest the claimed components for the same reasons discussed in Section VI.D.5. EX-2016, ¶152.

6. Claims 2-5

Because claims 2-5 depend from claim 1, they are valid under Ground 2B for at least the reasons discussed above for claim 1. EX-2016, ¶153.

F. GROUND 3: Claims 3-4

Because claims 3-4 depend from claim 1, they are valid under Ground 3 for at least the reasons discussed under Grounds 1A-2B above for claim 1. EX-2016, ¶154.

G. GROUND 4: Claim 5 is Not Obvious Over Grounds 1A-2B and Schuster

As discussed in Sections VI.B.3, VI.C.3, and VI.D.3, neither Shin nor Shimomura discloses or suggests “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” EX-1001, claim 5.

For Ground 4, the Petition alleges that claim 5 would have been obvious over Grounds 1A-2B and Schuster because Schuster discloses a “private communication mode ... when the user’s preceding voice interaction with the device is whispered.” Pet., 65-67 (citing EX-1016, Abstract, 1:7-11, 2:6-10, 2:59-67, 3:2-4, 3:12-18, 3:31-39, 6:34-36, 7:18-20). According to the Petition, because Schuster discloses distinguishing between a whispered voice and a non-whispered voice, a POSITA would have modified Shin and Shimomura to classify a voice based on whether it is a whisper. *Id.* The Petition’s proffered combinations of Grounds 1A-2B and Schuster are insufficient. EX-2016, ¶¶155-156.

Schuster discloses an unrelated process. It discloses shifting between modes (e.g., a “normal mode” and “private mode.”). EX-1016, 5:37-39; EX-2016, ¶157. The disclosed determination of whisper is not disclosed in the context of classifying voices. EX-2016, ¶157. The device does not respond to the whisper-determination. *Id.* It responds merely based upon mode: “Once the device has detected an intended private interaction, *this mode* change alters the way in which the device interacts with the user.” EX-1016, 3:31-33 (emphasis added); EX-2016, ¶157. The Petition does not establish if or how changing modes in a device relate to, or would improve, any of the other prior art. EX-2016, ¶157.

Regarding Schuster, Dr. Stern admitted that it does not provide “a lot of specificity” regarding how it would detect a whisper. EX-2017, 128:18-129:14. Dr.

Stern then attempted to gap fill Schuster's deficient disclosure with his own independent inferences. *Id.*

For Grounds 1A-2B, the Petition alleges that Shin and Shimomura disclose the claimed voice classification based on *distance ranges* calculated by a proximity sensor. *See* Pet., 10-14, 49-51. However, as the Petition acknowledges, "Schuster's device detects whispered input through *microphone signal analysis*." *Id.*, 65-66 (citing EX-1016, 3:2-4) (emphasis added). Schuster offers no guidance whatsoever on how a whisper could be detected based on distance at all—or why a POSITA would have been motivated to do so. EX-2016, ¶158. Quite the opposite.

Introducing distance would risk confounding the determination. *Id.*; EX-2017, 56:15-57:5. A POSITA would have recognized that Schuster's microphone signal analysis involves entirely different signals, metrics, and calculations than those in Shin and Shimomura's proximity sensor calculation. EX-2016, ¶158. A POSITA would not have known how to apply, and would have had no expectation of success in applying, Schuster's teaching of detecting whisper using a microphone to Shin and Shimomura's distance-based output schemes. *Id.* The Petition provides no articulated reasoning establishing that there would be an expectation of success or predictability. *Id.* To the extent the Petition argues that a POSITA would have been motivated to modify Shin and Shimomura such that different output content (as taught by Shin and Shimomura) is provided based on whether a whisper is

detected by a microphone (as taught by Schuster), such a modified Shin/Shimomura-Schuster device would not have met the limitations of claim 1, from which claim 5 depends. *See* EX-1001, claim 1 (“a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance ...”); EX-2016, ¶158.

H. GROUND 5: The Combination of Grounds 1-4 and Kristjansson Does Not Render the Challenged Claims Obvious

1. Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice

No combination of Grounds 1-4 and Kristjansson discloses the claimed voice classification for the reasons discussed above in Sections VI.B.1, VI.C.1, VI.D.1 and VI.E.1. EX-2016, ¶159.

2. Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Replacing Words in a First Output Sentence When the Voice is Determined to be the First Voice

No combination of Grounds 1-4 and Kristjansson discloses limitations the claimed word replacement for the reasons discussed above in Sections VI.B.2, VI.C.2, VI.D.2 and VI.E.2. EX-2016, ¶160.

3. Claim 5: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Determining the Voice Spoken by the User is a Whisper

No combination of Grounds 1-4 and Kristjansson discloses claim 5 for the reasons discussed above in Sections VI.B.3, VI.C.3, VI.D.3, and VI.E.3. EX-2016, ¶161.

4. Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Generation of a First and Second Output Sentence When the Voice has Been Classified as the First and Second Voice

No combination of Grounds 1-4 and Kristjansson discloses “generat[e/ing], when the voice is determined to be the first voice, a first output sentence, ... and to generat[ing], when the voice is not determined to be the second voice, a second output sentence” for the reasons discussed above in Sections VI.B.4, VI.C.4, VI.D.4, and VI.E.4. EX-2016, ¶162.

5. Claims 2-5

Because claims 2-5 depend from claim 1, they are valid under this ground for at least the reasons discussed above for claim 1. EX-2016, ¶163.

I. Other Deficiencies

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-2016, ¶¶164-165. 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

Petitioners fail to meet *Graham* factor two in its obviousness analysis for Grounds 1-5. 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*, 383 U.S. at 17-18). Petitioners are required to address all four *Graham* factors in its obviousness analysis. *WBIP, LLC v. Kohler Co.*, 829 F.3d 1317, 1328 (Fed. Cir. 2016). *Graham* factors one and two go hand-in-hand. POPR, 27-28. Petitioners do not explain what specifically is missing from the primary references, where it is found, or why a POSITA would have been motivated to bridge the gap. Petitioners do not sufficiently address *Graham* factor two at all.

Neither the Petition nor Petitioners' expert fails any better applying the other factors. They provide boilerplate recitations in support of factor one and then proceed to simply identify alternative disclosures for each limitation. One cannot merely point to disparate disclosures between multiple references and presume a POSITA would have been motivated to, or could with reasonable success, piece the references together in some unarticulated way to make it work.

The Petition only addresses *Graham* factor two in limited and isolated instances. For example, the Petition raises at least nine different potential combinations in Grounds 1A-5. And it proposes obviousness grounds to invalidate every limitation of every claim. However, none of the curt statements in the Petition are sufficient to meet the requirements of *Graham* factor 2. Pet., 14, 17-18, 23, 41, 52-53, 55, 61, 67. Thus, the Petition fails to establish a prima facie case of obviousness. POPR, 27-29.

3. Improperly Articulated Combinations for Ground 5

Ground 5 should be separated into at least six different Grounds. Ground 5 in the Petition states, "Claims 1-7 Would Have Been Obvious Over the Combinations in Grounds 1A-4 and Kristjansson." Pet., 67. It leaves open the possibility of at least ten combinations. Fairness dictates that the Board require that petitioners disclose their grounds with particularity and assess them as required under *KSR* and *Graham*. Petitioners have not done so.

VII. CONCLUSION

For all of the foregoing reasons, the Petition fails to demonstrate a reasonable likelihood that any of the challenged Claims 1-7 of the '675 patent are unpatentable.

Dated: December 24, 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 13,692 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 24, 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 24, 2025, on the following counsel of record for Petitioner:

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

Dated: December 24, 2025

Respectfully submitted,

By: / Steven J. Reynolds/

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