

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

DECLARATION OF JOSEPH C. MCALEXANDER III

TABLE OF CONTENTS

I.	BACKGROUND	1
II.	PROFESSIONAL QUALIFICATIONS	1
III.	MATERIAL CONSIDERED	2
IV.	LEGAL STANDARDS	3
A.	Enablement	4
B.	Written Description	4
C.	Anticipation	4
D.	Obviousness	5
E.	Claim Construction	7
V.	LEVEL OF ORDINARY SKILL IN THE ART	9
VI.	TECHNOLOGY BACKGROUND	9
VII.	THE '675 PATENT	10
A.	Specification	10
B.	Prosecution History	12
VIII.	CLAIM CONSTRUCTION	12
A.	“Classify the Voice into Either a First Voice or a Second Voice”	13
B.	“Whisper”	15
IX.	OVERVIEW OF THE ALLEGED PRIOR ART	16
A.	Shin (EX-1003)	16
B.	Shimomura (EX-1004)	17
C.	Aoyama (EX-1010)	18
D.	Iwase (EX-1017)	19

E.	Schuster (EX-1016).....	20
F.	Kristjansson (EX-1013).....	20
X.	CLAIMS	21
XI.	THE CHALLENGED CLAIMS OF THE '675 PATENT ARE VALID	24
A.	Deficiencies Present in All Grounds	24
1.	Equating Measurement to the Object Being Measured	24
2.	Failure to Address Obviousness Under Prevailing Precedent ..	27
B.	GROUND 1A: Shin in View of Aoyama Does Not Render the Challenged Claims Obvious.....	29
1.	Claims 1-7: Shin Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice.....	29
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	34
3.	Claim 5: Shin Does Not Disclose Determining the Voice Spoken by the User is a Whisper	39
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
C.	GROUND 1B: Shin in View of Iwase Does Not Render the Challenged Claims Obvious	43
1.	Claims 1-7: Shin Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice.....	43

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	43
3.	Claim 5: Shin Does Not Disclose Determining the Voice Spoken by the User is a Whisper	46
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	46
5.	Claims 1-7: Shin Does Not Disclose the Claimed Components Arranged as in the Claims.....	47
6.	Claims 2-5	47
D.	GROUND 2A: Shimomura in View of Aoyama Does Not Render the Challenged Claims Obvious.....	47
1.	Claims 1-7: Shimomura Discloses Measurement of Distance, Not Classification of Voices	47
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	50
3.	Claim 5: Shimomura Does Not Disclose Determining the Voice Spoken by the User is a Whisper	52
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	55
6.	Claims 2-5	56
E.	GROUND 2B: Shimomura in view of Iwase Does Not Render the Challenged Claims Obvious.....	56

1.	Claims 1-7: Shimomura Does Not Disclose Classification of a Voice as Either a First Voice or a Second Voice.....	56
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	56
3.	Claim 5: Shimomura Does Not Disclose Determining the Voice Spoken by the User is a Whisper	59
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	59
5.	Claims 1-7: Shimomura Does Not Disclose the Claimed Components Arranged as in the Claims	59
6.	Claims 2-5	60
F.	GROUND 3: Claims 3-4	60
G.	GROUND 4: Claim 5 is Not Obvious Over Grounds 1A-2B and Schuster	60
H.	GROUND 5: The Combination of Grounds 1-4 and Kristjansson Does Not Render the Challenged Claims Obvious	62
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.....	62
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	63
3.	Claim 5: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Determining the Voice Spoken by the User is a Whisper	63
4.	Claims 1-7: The Combination of Grounds 1-4 and Kristjansson Does Not Disclose Generation of First and Second Output	

	Sentence When the Voice has Been Classified as the First and Second Voice	63
5.	Claims 2-5	64
I.	Other Deficiencies	64
1.	Illogical Modifications	64
XII.	SUPPORT FOR MOTION TO AMEND	65
A.	The Substitute Claim Does Not Broaden the Claim Scope or Introduce New Subject Matter	65
1.	Written Description	65
2.	Enablement	70
XIII.	THE SUBSTITUTE CLAIM IS PATENTABLE OVER THE PRIOR ART RELIED UPON IN THE PETITION	71
A.	Shin and Shimomura Do Not Classify Voices as Described in the Substitute Claim	71
B.	Schuster Does Not Classify Voices As Described In The Substitute Claims	72

I have personal knowledge of the facts set forth in this declaration and, if called to testify as a witness, would testify under oath as follows:

I. BACKGROUND

1. I have been retained as an expert on behalf of SoundClear Technologies LLC (“SoundClear”) in the above-captioned matter (IPR2025-00673).

2. I have been asked by SoundClear to provide my expert opinion regarding the validity of claims 1-7 of U.S. Patent No. 11,244,675 (“the ’675 Patent”). For the reasons set forth below, it is my opinion that claims 1-7 of the ’675 Patent are valid.

3. I have also been asked to consider, and provide my opinions regarding Patent Owner’s Contingent Motion to Amend claim 5 of the challenged patent. In doing so, I discuss written description support and enablement for the amendment from the specification of the challenged patent.

II. PROFESSIONAL QUALIFICATIONS

4. I am a Registered Professional Engineer (#79454) in the State of Texas, and I am recognized as an inventor on 31 U.S. and a number of foreign patents.

5. I am President of M^εAlexander Sound, Inc., and the Managing Director of M^εAlexander Sound Pte Ltd. I have focused my expertise to support a number of clients in product, process, and operations analysis and investigation.

6. I have extensive experience with voice-control audio devices and audio processing in general including speech detection and spatial audio processing, beamforming, noise reduction, speech capture, near and far field detection, wireless audio transmission, audio device integration in automotive systems, spatial audio processing, audio system design and installation, and underwater acoustic signal detection.

7. Further details on various aspects of my professional experience and qualifications can be found in my curriculum vitae, which is provided as Exhibit 2014.

III. MATERIAL CONSIDERED

8. In preparing this declaration, I have reviewed the specification, claims and prosecution history of the '675 Patent.

9. I considered the following documents in preparing this Declaration.

Exhibit No.	Description
Petitioner (Amazon) Exhibits	
1001	U.S. Patent No. 11,244,675
1002	Declaration of Amazon's expert, Dr. Richard Stern
1003	U.S. Patent App. Publ. No. 2017/0083281 ("Shin")
1004	English Translation of Shimomura from Ex. 1005
1007	Excerpts from File History of U.S. Patent No. 11,244,675
1009	U.S. Patent App. Publ. No. 2017/0154626 ("Kim")
1010	U.S. Patent App. Publ. No. 2017/0337921 ("Aoyama")

Exhibit No.	Description
1011	U.S. Patent App. Publ. No. 2016/0284351 (“Ha”)
1012	U.S. Patent No. 9,489,172 (“Iyer”)
1013	U.S. Patent No. 10, 147,439 (“Kristjansson”)
1015	Nicolae Duta, Natural Language Understanding and Prediction: From Formal Grammars to Large Scale Machine Learning, 131 Fundamenta Informaticae 425 (2014) (“Duta”)
1016	U.S. Patent No. 9,680,983 (“Schuster”)
1017	U.S. Patent No. 11,183,167 (“Iwase”)
1018	U.S. Patent App. Publ. No. 2019/0103127 (“Tseretopoulos”)
Patent Owner (SoundClear) Exhibits	
Paper 8	Patent Owner’s Preliminary Response
2017	Transcript of Deposition of Petitioners’ Expert, Dr. Richard Stern
Other Documents	
Paper 1	Amazon’s Petition
Paper 15	Board’s Institution Decision
EX-2020	IPR2025-00565 Stern Deposition Transcript

IV. LEGAL STANDARDS

10. I am not an attorney, and I have not independently researched the law. SoundClear’s counsel has explained certain legal principles to me that I have relied on in forming my opinions set forth in this declaration. I have applied these legal principles in arriving at my opinions expressed in this declaration. They are set forth below.

11. I have been informed and understand the following principles.

A. Enablement

12. To fulfill the enablement requirement, the specification must enable the full scope of the invention as defined by its claims.

13. A patent may be invalid if the specification does not enable a person of ordinary skill in the art to make and use the claimed invention without undue experimentation. A person of ordinary skill in the art must be able to make the claimed invention across its full scope without undue experimentation.

B. Written Description

14. The written description must clearly allow a person of ordinary skill in the art to recognize that the inventor invented what is claimed and must reasonably convey to those skilled in the art that the inventor had possession of the claimed subject matter as of the filing date.

C. Anticipation

15. To anticipate a claim, a prior art reference must disclose each and every element of the claim. It must do so either explicitly or inherently.

16. To anticipate a claim, a prior art reference must also disclose those elements arranged as in the claim.

17. It is not proper to mix or combine incompatible embodiments in order to assert that a reference discloses each and every element of a claim.

D. Obviousness

An invention cannot be patented if the subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the invention. Also, while the prior art is compared to each claim on an element-by-element basis, the claimed invention as a whole must have been obvious to one of ordinary skill in order for a claim to be invalid.

18. The fundamental question in analyzing obviousness is whether, at the time of the invention, the subject matter of the claimed invention as a whole would have been obvious to a person of ordinary skill in the art to which the subject matter pertains, taking into account: (a) the scope and content of the prior art; (b) the differences between the prior art and the claims at issue; (c) the level of ordinary skill in the art; and (d) any objective indicia of non-obviousness referred to as secondary considerations. The secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism; (4) commercial success of a product covered by the patent; (5) industry praise of the claimed invention; and (6) copying of the invention.

19. A combination of prior art references must disclose, teach or suggest all of the recited limitations of the claim for the invention to be found obvious.

20. It is important in the obviousness inquiry to identify whether a reason existed at the time of the invention that would have motivated a person of ordinary

skill in the art in the relevant field to combine or modify the prior art references in the manner proposed by the petitioner so as to arrive at the claimed invention. Put another way, a finding of obviousness should be supported by an apparent reason to combine or modify the prior art references as proposed by the petitioner.

21. It is important in the obviousness inquiry to understand how the combination of references is supposed to work. An explanation of the operation of the combined references is often a prerequisite to showing that a person of ordinary skill in the art would have had an apparent reason to make the proposed combination and would have had a reasonable expectation of success.

22. A finding of obviousness requires the additional showing that a person of ordinary skill at the time of the invention would have selected and combined those prior art elements in the normal course of research and development to yield the claimed invention. In addition, a person of ordinary skill in the art should have had a reasonable expectation of success in combining the references as proposed.

23. In assessing obviousness, one must consider both the ordinary creativity and common sense of the person of ordinary skill in the art. It is impermissible for common sense to be applied to fill gaps in prior art that fails to teach or suggest a limitation of the claim.

24. Multiple references can be combined, with one another and/or with the knowledge of a person of ordinary skill in the art, in rendering a claim obvious.

However, obviousness cannot be established by simply demonstrating that each element was independently known in the prior art. Rather, it may be necessary to identify a reason, such as a teaching, suggestion, or motivation, which would have prompted a person of ordinary skill in the art to combine the elements in the way the claimed invention does.

25. Obviousness cannot be established through hindsight. That means the claimed invention cannot be used as a roadmap to combine elements from different pieces of prior art, or different embodiments of a single prior art reference, to create the claimed invention. The claimed invention as a whole must be compared to the prior art as a whole, and one must avoid aggregating pieces of prior art through hindsight that would not have been combined absent the patent inventor's insight.

E. Claim Construction

26. Claim construction is the process of determining the meaning of words used in a patent claim in order to understand the meaning of the claim. The proper construction of a term is how a person of ordinary skill in the art ("POSITA"), at the time of the invention, would have understood the term based on its use in the claims, specification, and prosecution history (the intrinsic record) of the patent.

27. The claims are not to be reviewed in this proceeding under the "broadest reasonable interpretation" standard.

28. Instead, the claims are to be reviewed under the *Phillips* standard in which the claims are to be interpreted based on their ordinary and customary meaning as understood by a person skilled in the art.

29. The claim construction inquiry begins and ends in all cases with the actual words of the claim. The language of the claim frames and ultimately resolves all issues of claim interpretation.

30. The inquiry into how a person of ordinary skill in the art understands a claim term provides an objective baseline from which to begin claim interpretation. That starting point is based on the well-settled understanding that inventors are typically persons skilled in the field of the invention and that patents are addressed to and intended to be read by others of skill in the pertinent art. Importantly, the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification.

31. A patent—including the specification and claims—and its prosecution history are considered “intrinsic evidence.” They provide aid in interpreting claim language in a patent. In reading the claim, it is improper to import limitations from the specification into the claim terms. It is also improper to narrow the scope of the claim terms by implicitly adding disclosed limitations that have no express basis in the claims.

32. Further, one must not provide a definition of claim terms divorced from the context of the patent and must not then turn to the specification only to look for grounds for limiting the claim scope.

33. The prosecution history of related patents and applications can also be relevant. Sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help interpret the claim language, but such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

V. LEVEL OF ORDINARY SKILL IN THE ART

34. 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.

VI. TECHNOLOGY BACKGROUND

35. I agree with Dr. Stern's recognition of the existence of voice-controlled devices in the 2010-era. Those included Alexa-enabled devices and Google's voice assistant. Each is centered on determining the content of speech, which is provided by a voice. They are not mere measurement devices.

VII. THE '675 PATENT

36. The '675 Patent was filed on March 7, 2019 as U.S. Patent Application No. 16/295,034. The '675 Patent also claims priority to Japanese Patent Application No. JP2018-044598 that was filed March 12, 2018. EX-1001, 1. Consequently, I understand that the priority date of the '675 Patent is at least as early as March 12, 2018.

A. Specification

37. 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. The focus is on user interaction with a device by speaking to it. More particularly, it relates to a device that analyzes a user’s voice. That voice includes multiple aspects, including the “content” contained in it. The '675 Patent disclosed device is not a mere measurement device. It is not simply a “volume measurement device” or “distance determination device.” The '675 Patent is focused on human voices and the analysis of different aspects of the human voice to classify voices not based on any singular characteristic, such as volume or distance.

38. As mentioned above, the '675 Patent describes “an output-content control device.” EX-1001, 1:44-45. The voice-content control device “compris[es]: a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second

voice.” *Id.*, 1:45-48. The patent also discloses an “output-content control method, comprising: acquiring a voice spoken by a user; analyzing the acquired voice to classify the acquired voice as either one of a first voice and a second voice.” *Id.*, 1:63-65.

39. According to the ‘675 Patent, one embodiment of the voice classifying unit “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.” *Id.*, 9:60-64. For example, the voice classifying unit can determine if a voice is a whisper or not and classify the voice accordingly. *See id.*, 9:65-10:4 (classifying the whisper as V1A). As another example, the voice classifying unit can perform Fourier-transformation on the time waveform to determine the intensity of the voice at different frequencies. *See id.*, 10:8-13 (“the voice classifying unit 42 acquires a time waveform of intensity of the voice V1, and performs Fourier-transformation on the time waveform, thereby acquiring a spectrum that indicates intensity of the voice V1 per frequency.”). That embodiment of the voice classifying unit can then classify the voice as a first voice V1A or second voice V1B by “using a peak frequency that is equal to or higher than a predetermined intensity in the spectrum as a feature value.” *Id.*, 10:12-16.

40. Other embodiments of the voice classifying unit use the peak frequency of the sound waves, the volume of the voice, the speaking speed of the voice, the

volume ratio between the speech and wind noise, the distance between the speaker and the device, or the mel-frequency cepstrum coefficient. *See id.*, 10:27-46 (explaining that the voice classifying unit sets certain feature values that distinguish between different voices). Further, “[t]he voice classifying unit 42 can use either one of the classification methods explained above, or can perform the classification by combining the classification methods explained above.” *Id.*, 11:1-4.

B. Prosecution History

41. I have reviewed the prosecution history of the ’675 Patent. EX-2014. Based on my review, I understand that the Examiner considered the same Aoyama and Schuster references that are cited in the Petition to support multiple grounds. Specifically, the Petition relies on Aoyama to support Grounds 1A, 2A, 3, 4, and 5. Petition, p. 5. The Petition relies on Schuster to support Grounds 4 and 5. *Id.*

42. The Petition states that the “Examiner erroneously allowed the claims.” Petition, p. 5. However, I note that the Petition does not provide any reasoning about why or how the Examiner erred in its analysis of Aoyama and Schuster.

VIII. CLAIM CONSTRUCTION

43. I have been informed that neither SoundClear nor Amazon proposed any terms for claim construction at the pre-institution stage of this proceeding. I also understand that the Board did not propose any terms for claim construction in its Institution Decision (“Decision”). As a POSITA, I opine that all claim terms should

be interpreted according to their ordinary and customary meaning. Below, I provide explanations of the ordinary and customary meaning for two claim terms central to the issues raised in this IPR proceeding.

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

44. 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*.”

45. Read in view of the specification, 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”). EX-1001, 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; EX-1001, 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.”)

46. In my opinion, 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. 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. A volume threshold that is consistent with both whisper and normal speech would not be fit for that purpose. 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.

47. 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.

B. “Whisper”

48. In my opinion, 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”:

Shin discloses that its device detects a range of speech volumes and rates. (Ex. 1003 ¶¶ [0037]-[0038].) Accordingly, a POSITA would have understood Shin’s disclosure to encompass receiving and processing as wide a range of speech as possible, including whispered speech. (Ex. 1002 ¶ 132.)

Petition, p. 38.

49. In my opinion, 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.”)).

50. 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.

IX. OVERVIEW OF THE ALLEGED PRIOR ART

A. Shin (EX-1003)

51. Shin (U.S. Patent App. Publ. No. 2017/0083281), entitled “Method and Electronic Device for Providing Content” was filed on September 19, 2016. *See* EX-1003. Shin claims priority to a Korean Patent Application No. 10-2015-0132488 filed on September 18, 2015. EX-1003 ¶ [0001].

52. Shin discloses an electronic device for providing content in response to user input. EX-1003, ¶ [0003]. Shin 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.*, ¶ [0010] (emphasis added). The Petition seems to agree: Shin operates “*based on* the user-to-device distance.” Petition, p. 12 (emphasis added).

B. Shimomura (EX-1004)

53. Shimomura (Japanese Unexamined Patent App. Publ. No. 2005/202076), entitled “Speech Control Device and Method and Robot Device” was filed on January 14, 2004. *See* EX-1004.

54. Shimomura is concerned only with distance and the volume output based on distance. Shimomura thus lacks all voice classification limitations, especially those based on an analysis of the voice and distance:

[Problem]

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

[Means for Solving]

By changing the speech style of the device or robot device when interacting with the user as necessary depending on the *distance* between the device or robot device and the user

EX-1004, at p. 1 (emphasis added); *id.*, ¶ [0012] (“the speech form of the device during a dialogue with the user is changed *as necessary according to the distance* between the device and the user.”) (emphasis added)).

55. Shimomura discloses a “change[] in intonation.” EX-1004, ¶ [0092] (“changes in intonation (emphasis at the end of each phrase) ...”).

C. Aoyama (EX-1010)

56. Aoyama et al. (U.S. Patent App. Publ. No. 2017/0337921 A1) entitled “Information Processing Device, Information Processing Method, and Program” was filed on November 26, 2015.

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

58. 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 assess 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.

59. Second, Aoyama separately discloses “control[ing] an amount of information output as the response....” *Id.*, ¶ [0145]. It does so based on “acquired information indication 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].

60. I note that 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.

D. Iwase (EX-1017)

61. Iwase et al. (U.S. Patent No. 11,183,167), entitled “Determining an Output Position of a Subject in a Notification Based on Attention Acquisition Difficulty” was filed on December 25, 2017.

62. 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.

63. 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.

64. 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.

E. Schuster (EX-1016)

65. Schuster et al. (U.S. Patent No. 9,680,983), entitled “Privacy Mode Detection and Response Over Voice Activated Interface” was filed on June 16, 2016.

66. 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.

67. I note that the USPTO issued the ’675 Patent over Schuster. During prosecution, the Examiner considered Schuster and specifically allowed the challenged claims over Schuster.

F. Kristjansson (EX-1013)

68. Kristjansson et al. (U.S. Patent No. 10,147,439), entitled “Volume Adjustment for Listening Environment” was filed on March 30, 2017. *See* EX-1013.

69. Kristjansson discloses a device that uses distance to determine sound pressure where the user is located. EX-1013, Abstract.

X. CLAIMS

70. The following claims appear in the '675 Patent.

Claims
1. An output-content control device comprising: [a] a voice acquiring unit configured to acquire a voice spoken by a user; [b] a voice classifying unit configured to calculate a distance between the user and the output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance; [c] an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user; [d] a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and [e] an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information, wherein [f] the output-content generating unit is further configured to generate, 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, and [g] to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information

Claims
and the content information.
2. The output-content control device according to claim 1, wherein [a] the notification information includes type information indicating a type of the content information, and [b] the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be placed with the other word based on a word included in the content information and the type information that is associated with the content information.
3. The output-content control device according to claim 2, wherein the output-content generating unit is further configured to read, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.
4. The output-content control device according to claim 3, wherein [a] the relationship information further includes information of type information indicating a type of the first word, and; [b] the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word
5. The output-content control device according to claim 1, wherein the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.

Claims
6. An output-content control method comprising: [a] acquiring a voice spoken by a user; [b] calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance; [c] analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user; [d] acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and [e] generating, 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; and [f] generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.
7. A non-transitory storage medium that stores an output-content control program that causes a computer to execute: [a] acquiring a voice spoken by a user; [b] calculating a distance between the user and an output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance; [c] analyzing the acquired voice to detect intention information indicating what kind of information is wished to be acquired by the user; [d] acquiring notification information which includes content information as a content of information to be notified to the user based on the intention information; and

Claims
[e] generating, 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; and
[f] generating, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.

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

A. Deficiencies Present in All Grounds

1. Equating Measurement to the Object Being Measured

71. In my opinion, all of the grounds in the Petition equate the concept of an objection 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, 95:12-17 (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.*, 119:17-19.

72. For example, 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; *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. None of the prior art discloses classifying a voice separate from measuring a characteristic of the voice.

73. 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”).

74. 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.

EX-2017, 91:13-18.

75. In my opinion, 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. EX-2017, 56:15-57:5 (“I don’t know.”). In his view, that is not relevant. His conclusion is wrong.

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

77. There is no indication that a POSITA would have been motivated to change how Shin’s purported classification 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).

2. Failure to Address Obviousness Under Prevailing Precedent

78. The Board provides two examples of obviousness rationales (shown below). Decision, p. 26-27.

the proposed combination is merely the addition of one known element (i.e., “Aoyama’s natural language	the proposed combination also represents the addition of one known element (i.e., “Aoyama’s word
--	--

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

79. In my opinion, both are incomplete. 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.

80. Further, the first rationale does not address any actual difference between the claims and the prior art. 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).

81. 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. 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.

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

82. As discussed 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. Petition, pp. 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 to the user based on the user-to-device distance.” Petition, p. 22. 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.” Petition, p. 12 ((citing EX-1002, ¶60) (emphasis added) (citing only expert declaration)).

83. 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.” Petition, p. 12. That is, garden variety natural language processing to determine user intent. 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 or the ’675 Patent. Neither the Petition nor Petitioners’ expert provides any reasoned analysis supporting the conclusion that Shin classifies voices.

84. 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. Petition, p.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)

85. EX-1003, Fig. 3.

86. 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. Nor can it expressly disclose an analysis of a "voice," which is not at all synonymous with "distance."

87. One example is sufficient to demonstrate the failed reasoning here. Assume two users, with two different voices, were standing the same distance from the device disclosed in Shin. Table 3 shows that Shin would not distinguish between the voices of those users. The only information determined by Shin would be distance. And it would be the same for each voice.

88. 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]. And Shin confirms the distinction between “voice” and “distance.” *Id.* Shin states that “processor 120 may determine a distance between the user and the electronic device 101 based on at least one of the *volume* of the user’s speech obtained through an analysis of a *voice* input or the *distance* computed, calculated, or measured by the distance detection module 180.” *Id.*, ¶ [0053]. Thus, a POSITA would never have understood Shin’s disclosure of distance to be synonymous with “voice.” Nor would a POSITA have understood Shin’s determination of a distance to be synonymous with classification.

89. 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. Petition, p. 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.

90. Crucially, Shin’s Table 3 does not even disclose voice classification under Petitioners’ expert’s own understanding. Specifically, Petitioners’ expert 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.

91. 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.

92. In my opinion, 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).

93. 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.

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

94. 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. In my opinion, it does not.

95. The Petition alleges that Shin discloses user-to-device distance ranges at which the output content is abbreviated. Petition, pp. 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. Shin provides no indication that the word “abbreviation” means or includes “replacing words.”

96. 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.

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

98. The Petition states that replacing words would have been obvious over Aoyama. Petition, pp. 23-25. Specifically, the Petition states that “Aoyama’s device tailors output by replacing words based on the user or situation.” Petition, p. 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]. 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. 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.” *Id.*, ¶¶ [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. Nor does it provide any basis for a word replacement based on a voice classification.

99. 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]. Aoyama teaches that those features involve separate and distinct processes and metrics. 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. Aoyama teaches that words are *only* to be chosen in response to “speaking style.” EX-1010, ¶ [0049]. That “speaking style” is tied to the specific words a user tends to use. 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. Indeed, doing so would have defeated the purpose of Aoyama. It would have led to a response with which “the user may feel uncomfortable.” EX-1010, ¶ [0049]. For example, combining Shin and Aoyama’s 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. 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.

100. In my opinion, 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. Like Shin, Aoyama discloses changing the “amount of information output as the response.” EX-1010, ¶¶ [0145]-[0147]. 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.* 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 Shin's information-modifying portion. And, based on Aoyama's teachings, that would have remained distinct from word replacement.

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

101. 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.

102. 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." Petition, p. 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.

103. The Petition asserts that a POSITA would have understood a whisper to be a subset of "a range of speech volumes and rates." Petition, p. 38. In my opinion, there is no basis for equating a whisper to be a range of volumes and rates.

At best, that position conflates the object being classified with the manner of classifying it. Further, Petitioners' approach would vitiate any difference in scope between claim 1 and 5. 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.

104. In my opinion, 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. It would not be configured as claimed.

105. I note that 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. Decision, 22 n.5. I understand that a reference can be cited to establish background of knowledge but not to modify the references in the combination. Thus, Ground 1A fails for that reason alone. It does not provide Iyer as context. It provides Iyer to supply the missing limitation of determining a "whisper."

106. The Petition does not identify any express or inherent disclosure in Shin that allegedly meets the elements of claim 5. 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]*.” Petition, p. 38. But claim 5 requires more than the capacity to receive and process whispered speech. It requires a determination that “the voice spoken by the user is a whisper.” *See also* EX-1001, 9:58-10:8. 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). 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

107. 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.

108. For reasons discussed above, Shin does not disclose or suggest classifying the voice as the first and second voice as claimed. 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. Aoyama does not cure the deficiencies of Shin. Therefore Ground 1A fails to invalidate any claim.

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

109. For Ground 1A, the Petition focuses on the claimed functions without identifying what component allegedly accomplishes those functions. 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. Petition, p. 67 (addressing “intention analyzing unit” and “notification-information acquiring unit”). Amazon’s attempt

to cure the deficiencies of Ground 1A is cursory. The Petition spends most of its pages repeating incomplete obviousness rationales. Petition, pp. 67-69.

110. 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.

6. Claims 2-5

111. 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.

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

112. Shin does not disclose the claimed voice classification for the same reasons discussed in above. Iwase does not cure the deficiencies of Shin on this claim limitation, and the Petition does not rely on Iwase to disclose it. 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

113. 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 above, in my opinion, Shin does not disclose it.

114. 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.’” Petition, pp. 41-42 (citing EX-1017, 1:54-2:18, 4:17-29, 6:32-35, 12:26-42, 15:6-40). But Iwase does not disclose replacing words based on voice classification.

115. 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.

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

117. In my opinion, a POSITA would have no reason to combine Shin and Iwase in the manner claimed. Iwase is not concerned at all with distance. It does not improve any process that incorporates distance. The Petition provides no basis for asserting that a POSITA would have been motivated to make the combination. The Petition’s sole rationale is that Iwase’s feature would allow a user “to grasp notification contents more effectively.” Petition, p. 42 (citing 2:11-14, 15:24-28, 21:66-22:8). That is an empty, conclusory assertion. 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. It reduces to purely circular reasoning: that a POSITA would have combined Shin and Iwase to arrive at the combination. Petition, pp. 42-44. 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. 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. 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

118. In my opinion, the combination of Shin and Iwase does not disclose or suggest claim 5 for the same reasons discussed above. 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.

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

119. 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. In my opinion, for reasons discussed in above, 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.

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

120. I addressed Shin and Iwase above. That discussion applies to the combination of both references here. In my opinion, the combination of Shin and Iwase does not disclose or suggest the claimed components for the same reasons discussed above.

6. Claims 2-5

121. Claims 2-5 depend from claim 1. In my opinion, they are valid under Ground 1B for the same reasons discussed above for claim 1.

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

122. As I discussed 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.

123. Shimomura does not disclose any manner of classifying voices. The Petition acknowledges that Shimomura discloses a “device” that “controls speech forms ... according to the physical distance to the conversation partner.” Petition, p. 47 (citations omitted). The only basis the Petition provides for “classification” of a “voice” is in connection with Fig. 7. The Petition does not identify any express

disclosure of either aspect. Indeed, the Petition does not cite to a single disclosure of a “voice.” All mentions of “voice” (including those with quotes around them) are taken from Petitioners’ expert or the ’675 Patent. Neither the Petition nor Dr. Stern provides any reasoned analysis supporting the conclusion that Shimomura classifies voices. Thus, in my opinion, the Petition fails to demonstrate that Shimomura discloses voice classification.

124. In my opinion, 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.”

125. Similarly, to the extent that Shimomura discloses any required, claimed classification, it discloses classes of distances, not voices. 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

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

126. Dr. Stern added the “First Voice” and “Second Voice” labels to Fig. 7. In my opinion, annotations aside, Figure 7 discloses one thing: a calculation of

distance and nothing else. That determination of distance cannot expressly disclose a classification, because it is a measurement of a continuous variable, not a discrete classification. Nor can it expressly disclose an analysis of a “voice,” which is not at all synonymous with “distance.”

127. Shimomura confirms the distinction between “voice,” “distance,” and “volume.” 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).

128. In my opinion, a POSITA would never have understood Shimomura’s disclosure of distance to be synonymous with “voice.” Nor would a POSITA have understood Shimomura’s determination of a distance to be synonymous with a classification.

129. 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. 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

130. I discussed both Shimomura and Aoyama above. In my opinion, the combination of both references together does not disclose replacing words based on voice classification for similar reasons discussed above. The Petition does not assert that Shimomura discloses word replacement. It relies on Aoyama for that limitation. Petition, p. 55. And the proposed combination of Shimomura with Aoyama fails for the same reasons.

131. 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]. Aoyama teaches that those features involve separate and distinct processes and metrics. 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. Aoyama teaches that words are *only* to be chosen in response to "speaking style." *Id.*, ¶ [0049]. That "speaking style" is tied to the specific words a user tends to use. 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. Indeed, doing so would have defeated the purpose of Aoyama. 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. 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.

132. In my opinion, 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. Like Shimomura, Aoyama discloses changing the “amount of information output as the response.” EX-1010, ¶¶ [0145]-[0147]. 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.* 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.

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

133. 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.

134. The Petition states 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.” Petition, p. 58. The latter statement does not support any conclusion that processing speech includes determining whether speech input was provided by a whisper. Thus, as discussed similarly above for Shin, there is no stated basis upon which one would conclude that Shimomura discloses any manner of whisper determination. 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.

135. 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. It would not be configured as claimed. Dr. Stern’s application of Shimomura fails for the same reason as discussed in connection with Shin. *Id.*, 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.”).

136. The Petition further alleges that “[e]ven 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.” Petition, p. 58. The Petition does not provide any support for that assertion. The Petition improperly relies on prior art not included in the proffered combination (Iyer), as discussed above. Thus, the assertion is at best conclusory.

137. 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. 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.” Petition, p. 58. But claim 5 requires more than the capacity to receive and process whispered speech. It requires a determination that “the voice spoken by the user is a whisper.” 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).

138. 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). 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

139. 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.

140. I discussed Shimomura above. 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.

141. 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.

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

142. For Ground 2A, the Petition focuses on the claimed functions without identifying what component allegedly accomplishes those functions. In my opinion,

Ground 2A does not address the details or arrangement of the components/steps of any claim of the '675 Patent. 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. Petition, p. 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. *See* Petition, pp. 67-69.

6. Claims 2-5

143. Claims 2-5 depend from claim 1. In my opinion, they are valid under Ground 2A for the same reasons discussed above for claim 1.

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

144. Shimomura does not disclose the claimed voice classification for the same reasons I discussed above.

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

145. I discussed Iwase and Shimomura above. The combination of Shimomura and Iwase does not disclose replacing words based on voice

classification for similar reasons discussed above. In my opinion, the Petition does not assert that Shimomura discloses word replacement. It relies on Iwase for that limitation. Petition, pp. 60-62. But the combination of Shimomura with Iwase still fails to disclose the claim limitation.

146. 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.

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

148. In my opinion, a POSITA would not have any 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. The Petition provides no basis for asserting that a POSITA would have been motivated to make the combination. The Petition's sole rationale is that Iwase's feature would allow a user "to grasp notification contents more effectively." Petition, p. 42 (citing EX-1017, 2:11-14, 15:24-28, 21:66-22:8). That is an empty, conclusory assertion. 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. It reduces to purely circular reasoning: that a POSITA would have combined Shimomura and Iwase to arrive at the combination. Petition, pp. 42-44. 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. 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.

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

149. I discussed Shimomura and Iwase above. In my opinion, the combination of both Shimomura and Iwase does not disclose or suggest claim 5 for the same reasons discussed above.

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

150. 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.

151. I discussed Shimomura and Iwase above. In my opinion, the combination of both Shimomura and Iwase does not disclose or suggest this limitation for the same reasons discussed above. 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.

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

152. I discussed Shimomura and Iwase above. In my opinion, the combination of both Shimomura and Iwase does not disclose or suggest the claimed components for the same reasons discussed above.

6. Claims 2-5

153. Claims 2-5 depend from claim 1. In my opinion, they are valid under Ground 2B for the same reasons discussed above for claim 1.

F. GROUND 3: Claims 3-4

154. In my opinion, 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.

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

155. 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.

156. In Ground 4, the Petition states 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.” Petition, pp. 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.

157. Schuster discloses an unrelated process. It discloses shifting between modes (*e.g.*, a “normal mode” and “private mode.”). EX-1016, 5:37-39. The disclosed determination of whisper is not disclosed in the context of classifying voices. The device does not respond to the whisper-determination. 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.” *Id.*, 3:31-33 (emphasis added). The Petition does not establish if or how changing modes in a device relate to, or would improve, any of the other prior art.

158. For Grounds 1A-2B, the Petition alleges that Shin and Shimomura discloses the claimed voice classification based on distance ranges calculated by a proximity sensor. *See* Petition, pp. 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). Schuster offers no guidance whatsoever on how a whisper could be detected based on distance at all—or why a POSITA would have been motivated to do so. Quite the opposite. Introducing distance would risk confounding the determination. EX-2017, 56:15-57:5 A POSITA would have recognized that Schuster’s microphone signal analysis involves entirely different signals, metrics, and calculations than those in Shin and Shimomura’s proximity sensor calculation. *See, e.g.*, EX-1016, 3:1-11, 6:30-41. 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. The Petition provides no articulated reasoning establishing that there would be an expectation of success or predictability. 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").

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

159. I discussed Grounds 1-4 above. In my opinion, the combination of Grounds 1-4 with Kristjansson does not disclose the claimed voice classification for the same reasons discussed above.

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

160. I discussed Grounds 1-4 above. In my opinion, the combination of Grounds 1-4 with Kristjansson does not disclose the claimed word replacement for the same reasons discussed above.

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

161. I discussed Grounds 1-4 above. In my opinion, the combination of Grounds 1-4 with Kristjansson does not disclose claim 5 for the same reasons discussed above.

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

162. I discussed Grounds 1-4 above. In my opinion, the combination of Grounds 1-4 with Kristjansson does not disclose “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.

5. Claims 2-5

163. Claims 2-5 depend from claim 1. In my opinion, they are valid under this Ground for the same reasons discussed above for claim 1.

I. Other Deficiencies

1. Illogical Modifications

164. 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.

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. They purposely chose simpler, more discrete approaches. 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. It would have mandated a more complex approach than implementing a single-variable measurement in simple, two-dimensional tables.

XII. SUPPORT FOR MOTION TO AMEND

166. I have been informed that SoundClear is seeking to amend claim 5 of the '675 Patent contingently in the form of proposed substitute claim 8. As a POSITA, I opine herein on the proposed amendment.

A. The Substitute Claim Does Not Broaden the Claim Scope or Introduce New Subject Matter

167. The specification provides sufficient enablement and written description support for substitute claim 8, as shown in the chart below. A person of ordinary skill in the art would have understood that the inventor possessed substitute claim 8 as of the priority date, March 12, 2018. *See* EX-1001.

1. Written Description

168. In my opinion, according to the specification, the voice classification unit classifies the first voice as whisper as a result of analysis. EX-1001, 9:65-10:1. The voice classifying unit classifies the voice as a second voice when it is determined not to be a whisper. EX-1001, 10:1-4. A whisper is defined as “an unvoiced voice spoken without vibrating the vocal chords or one in which a peak frequency is equal to or lower than a threshold.” EX-1001, 10:4-8. The voice classifying unit performs the classification of a voice as a whisper or not a whisper. EX-1001, 10:47-53.

169. Those portions of the specification cited below provide support for the proposed amended claim language. They also demonstrate that the inventor possessed the full invention, as amended, as of the priority date, March 12, 2018.

Proposed Claim 8	Exemplary Support in the '675 Patent for the Amended Language (EX-1001)
8. An output-content control device comprising:	<i>See Original Claim Language Table below.</i>
a voice acquiring unit configured to acquire a voice spoken by a user;	<i>See Original Claim Language Table below.</i>
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;	<i>See Original Claim Language Table below.</i>
an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;	<i>See Original Claim Language Table below.</i>
a notification-information acquiring unit configured to acquire notification	<i>See Original Claim Language Table below.</i>

Proposed Claim 8	Exemplary Support in the '675 Patent for the Amended Language (EX-1001)
information which includes content information as a content information to be notified to the user based on the intention information; [[and]]	
an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information, wherein the output-content generating unit is further configured to generate, 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, and to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information; and	See Original Claim Language Table below.
the voice classifying unit is further configured to determine the voice to be [[the]] a predetermined voice when the voice spoken by the user is <u>determined to be a</u> whisper <u>based upon whether a peak frequency is equal to or lower than a threshold.</u>	9:58-10:67; 10:4-8, 10:17-22, 10:48-54, 17:44-67; 19:23-30.

170. In my opinion, the following sections of the specification provide support for the asserted claim language in claims 1 and 5.

Original Claim Language	Exemplary Support in the '675 Patent for the Original Language (EX-1001)
1. An output-content control device comprising:	2:64-3:14; 3:10-56; 4:15-24; 4:58-5:3; 5:14-30; 6:54-56; 9:14-41; 10:32-38; 14:44-15:3; 17:21-67; 18:22-27; 18:23-64; 19:23-45. Figs. 1-2.
a voice acquiring unit configured to acquire a voice spoken by a user;	2:66-3:4; 3:61-67; 4:1-5; 9:58-64; 16:45-48; 17:25-32. Figs. 1-2, and 9.
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;	3:61-67; 9:58-11:4; 11:5-8; 11:53-56; 16:64-67; 17:21-28; 19:23-25. Figs. 2 and 9.
an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user;	3:55-67; 4:15-5:3; 5:50-6:23; 6:49-54; 7:31-65; 9:14-41; 11:5-21; 16: 41-57; 17:21-32. Figs. 2-3, and 9.
a notification-information acquiring unit	3:55-67; 5:4-13; 9:1-56; 17:21-43.

Original Claim Language	Exemplary Support in the '675 Patent for the Original Language (EX-1001)
configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information; and	Fig. 2.
an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information, wherein the output-content generating unit is further configured to generate, 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, and to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.	3:55-67; 6:56-7:56; 8:48-67; 9:42-57; 11:53-16:30; 17:1-20; 18:1-19:22; 20:42-49. Figs. 2, 4-6, and 9.
5. The output-content control device according to claim 1, wherein	2:64-3:14; 3:10-56; 4:15-24; 4:58-5:3; 5:14-30; 6:54-56; 9:14-41; 10:32-38; 14:44-15:3; 17:21-67; 18:22-27; 18:23-64; 19:23-45. Figs. 1-2.
the voice classifying unit is further configured to determine the voice to	9:64-10:22; 17:44-67; 19:23-30.

Original Claim Language	Exemplary Support in the '675 Patent for the Original Language (EX-1001)
be the predetermined voice when the voice spoken by the user is a whisper.	

2. Enablement

171. In my opinion, the additional detail of claim 8 would satisfy the enablement requirement. The additional elements are sufficiently enabled by the specification in view of the factors set forth for enablement. The specification describes how the voice classifying unit of the output-content control device determines a voice to be a whisper or not a whisper. Claim 8 would limit the whisper to one that is determined based on “whether a peak frequency is equal to or lower than a threshold.” Motion to Amend, Appx. A. The specification provides an example to explain how that can be accomplished. The voice acquiring unit receives voice input. EX-1001, 9:58-60. The voice classifying unit determines if a voice is a whisper or not a whisper. *Id.*, 9:65-10:8. It does so by, for example, performing a Fourier-transformation on the time waveform. *Id.*, 10:9-12. The specification provides more detail as to how the Fourier transformation can be used. It can be used to produce a spectrum indicating the intensity of the voice input. *Id.*, 10:9-12. The voice classifying unit then analyzes a peak frequency in that spectrum. *Id.*, 10:9-12.

It then determines if the peak frequency is equal to or lower than a predetermined intensity. *Id.*, 10:12-16. The voice classifying unit then determines the voice to be a whisper when the peak frequency is equal to or lower than a threshold value. *Id.*, 10:16-24. If the peak frequency is “larger” than a threshold value, the voice is determined to not be a whisper. *Id.*

XIII. THE SUBSTITUTE CLAIM IS PATENTABLE OVER THE PRIOR ART RELIED UPON IN THE PETITION

A. Shin and Shimomura Do Not Classify Voices as Described in the Substitute Claim

172. In my opinion, Shin does not disclose any attempt to identify whether a voice is a whisper or not a whisper, let alone in connection with any voice classification. *See* EX-1003. It is only concerned with distance. The Petition asserts that “a POSITA would have understood Shin’s disclosure to encompass receiving and processing as wide a range of speech as possible including whispered speech.” Petition, p. 38 (citing EX-1002, ¶132). But it acknowledges that Shin’s device “did not explicitly support processing whispered speech.” *Id.*, (citing EX-1002, ¶¶133-37). And it provides no basis for the statements about what a POSITA would have understood.

173. Substitute claim 8 would be patentable over the prior art. Shin does not disclose that a whisper is determined based on whether a peak frequency is equal to or lower than a threshold. *See* Motion to Amend, Appx. A. Substitute claim 8 would

overcome Shimomura for the same reasons. The “speech synthesis unit” described in Shimomura does not disclose a processor that would be able to determine a whisper, let alone based on whether a peak frequency is equal to or lower than a threshold. *See* EX-1004.

B. Schuster Does Not Classify Voices As Described In The Substitute Claims

174. In my opinion, the closest applicable disclosure in the prior art simply mentions whisper. The Petition asserts that Schuster’s device can detect “whispered input.” Petition, pp. 65-66 (citing EX-1016, 3:2-4). But it does so in a completely different context. The Examiner expressly considered Schuster and allowed claim 5 over Schuster. EX-2014, 133-134, 179-190. Proposed claim 8 would overcome Schuster for other reasons as well. It defines the term “whisper” based upon the manner in which it is determined to be one, and recites a specific manner of doing so. That is, whether a peak frequency is equal to or lower than a threshold. Motion to Amend, Appx. A. Schuster does not address a peak frequency being equal to or lower than a threshold when describing whispered speech. EX-1016, 3:4-7 (defining “whisper” as “unvoiced or substantially unvoiced speech as well as hushed speech”).

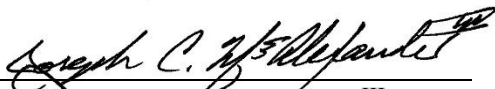
175. Schuster’s high-level disclosure about using “cepstral analysis ... to detect whispered input” (EX-1016, 3:8-11) is at best insufficiently specific. “Cepstral analysis” broadly defines a method of capturing audio. A POSITA would

not have understood that reference to cover any particular manner of analyzing audio, certainly not a specific manner of detecting whispers. Schuster does not disclose that it involves determining “whether a peak frequency is equal to or lower than a threshold.” In fact, Schuster does not provide any explanation for how it applies the cepstral analysis to the incoming microphone signal, let alone how it is used to differentiate a whisper from not a whisper. *See, e.g.*, EX-1016, 3:10-11, 6:38-41.

176. Further, in view of claim 1, Schuster does not mention any distinguishing between multiple voices. Therefore, substitute claim 8 overcomes all the prior art. Neither Schuster nor any other prior art reference discloses determining whether “a peak frequency is equal to or lower than a threshold.” *See* EX-1016.

I, Joseph C. McAlexander ^{III}, P.E., do hereby declare and state, that all statements are made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I am over the age of 21 and am competent to make this declaration. These statements were made with the knowledge that willful false statements are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: December 24, 2025


Joseph C. McAlexander ^{III}, P.E.