

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

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

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner.

IPR2025-00565
Patent 11,069,337 B2

Before MICHAEL R. ZECHER, DAVID C. McKONE, and
IFTIKHAR AHMED, *Administrative Patent Judges*.

ZECHER, *Administrative Patent Judge*.

JUDGMENT
Final Written Decision
Determining All Challenged Claims Unpatentable
35 U.S.C. § 318(a)

I. INTRODUCTION

Petitioner, Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively, “Amazon”), filed a Petition requesting an *inter partes* review (“IPR”) of claims 1–5 of U.S. Patent No. 11,069,337 B2 (Ex. 1001, “the ’337 patent”). Paper 1 (“Pet.”). Patent Owner, SoundClear Technologies LLC (“SoundClear”), filed a Preliminary Response. Paper 9 (“Prelim. Resp.”). Taking into account the arguments presented in SoundClear’s Preliminary Response, the Board determined that the information presented in the Petition established that there was a reasonable likelihood that Amazon would prevail with respect to challenging at least one of claims 1–5 of the ’337 patent. Pursuant to 35 U.S.C. § 314, the Board instituted this IPR on August 27, 2025, as to claims 1–5 of the ’337 patent based on all grounds raised in the Petition. Paper 12 (“Dec. on Inst.”).

During trial, SoundClear filed a Patent Owner Response (Paper 23, “PO Resp.”), Amazon filed a Reply (Paper 27, “Pet. Reply”), and SoundClear filed a Sur-reply (Paper 34, “PO Sur-reply”). An oral argument was held on May 20, 2026, and the transcript of the hearing is included in the record. Paper 38 (“Tr.”).

We have jurisdiction under 35 U.S.C. § 6. This decision is a Final Written Decision under 35 U.S.C. § 318(a) as to the patentability of claims 1–5 of the ’337 patent. For the reasons we identify below, we hold that Amazon has demonstrated by a preponderance of the evidence that these claims are unpatentable.

A. Real Party in Interest (“RPI”)

Amazon identifies Amazon.com, Inc., Amazon.com Services, LLC, and Amazon Web Services, Inc. as RPIs. Pet. 72; Paper 16 (Amazon’s Updated Mandatory Notices), 1. SoundClear identifies itself, which “is a wholly owned subsidiary of SoundClear LLC,” as an RPI. Paper 33 (SoundClear’s Updated Mandatory Notices), 1.

B. Related Matters

The parties indicate that the ’337 patent is the subject of the following two district court cases: (1) *SoundClear Techs. LLC v. Amazon.com Inc.*, No. 1:24-cv-01283 (E.D. Va. filed July 25, 2024); and (2) *SoundClear Techs. LLC v. Google LLC*, No. 3:24-cv-00540 (E.D. Va. filed July 25, 2024). Paper 16 at 1; Paper 33 at 2. In addition to this Petition, Google LLC filed the following two petitions challenging the ’337 patent:

(1) *Google LLC v. SoundClear Techs. LLC*, IPR2025-01123, Paper 1 (PTAB June 12, 2025) (challenging claims 1–5 of the ’337 patent); and (2) *Google LLC v. SoundClear Techs. LLC*, IPR2025-01597, Paper 1 (PTAB Sept. 29, 2025) (challenging claims 1–5 of the ’337 patent). Paper 16 at 1; Paper 33 at 2. Amazon also filed the following five petitions challenging related patents owned by SoundClear: (1) *Amazon.com, Inc. v. SoundClear Techs. LLC*, IPR2025-00673, Paper 1 (PTAB Feb. 27, 2025) (challenging claims 1–7 of U.S. Patent No. 11,244,675 B2); (2) *Amazon.com, Inc. v. SoundClear Techs. LLC*, IPR2025-01067, Paper 1 (PTAB May 28, 2025) (challenging claims 1–15 of U.S. Patent No. 9,070,374 B2); (3) *Amazon.com, Inc. v. SoundClear Techs. LLC*, IPR2025-01080, Paper 1 (PTAB May 30, 2025) (challenging claims 1–5, 7, and 8 of U.S. Patent No. 9,804,819 B2); (4) *Amazon.com, Inc. v. SoundClear Techs. LLC*, IPR2025-01096, Paper 1

(PTAB June 2, 2025) (challenging claims 1–20 of U.S. Patent No. 9,031,259 B2); and (5) *Amazon.com Inc. v. SoundClear Techs. LLC*, IPR2025-01368, Paper 1 (PTAB July 30, 2025) (challenging claims 1–3 and 10–12 of U.S. Patent No. 9,223,487 B2). Paper 16 at 1.

C. The '337 Patent

The '337 patent, titled “Voice-Content Control Device, Voice-Content Control Method, and Non-Transitory Storage Medium,” issued from U.S. Patent Application No. 16/290,983, filed on March 4, 2019. Ex. 1001, codes (54), (21), (22). The '337 patent claims priority to Japanese Patent Application No. JP2018-039754, filed on March 6, 2018. *Id.* at code (30).

The '337 patent generally relates to “a voice-content control device, a voice-content control method, and a non-transitory storage medium” that stores a program that causes a computer to execute certain steps. Ex. 1001, 1:15–17. The '337 patent acknowledges that the prior art discloses “a voice control device that analyzes [the] detected voice of a user and performs processing according to the user’s intention.” *Id.* at 1:21–24. The prior art also discloses “a voice control device, which outputs, via voice, . . . processing intended by the user . . . , or . . . voice content of a user’s inquiry.” *Id.* at 1:24–28. The '337 patent, however, states that, “when a voice processing device that outputs voice is used, the output voice may be heard by a person who is not the user of the voice processing device and is around the voice processing device.” *Id.* at 1:29–32. According to the '337 patent, it would be desirable to suppress the “influence of the output voice to people other than the user” so that “content of the output voice to the user is . . . made adequately understandable.” *Id.* at 1:39–42.

The '337 patent addresses this need by providing a voice-content control device, a voice-content control method, and a non-transitory storage medium that stores a program that causes a computer to execute certain steps, which are capable of acquiring a voice, classifying the voice, analyzing the voice, acquiring notification information, and generating output based on the notification information. *See, e.g.*, Ex. 1001, 1:48–64 (disclosing “a voice-content control device”), 1:65–2:11 (disclosing “a voice-content control method”), 2:12–27 (disclosing “a non-transitory storage medium” that stores a program that causes a computer to execute certain steps). Figure 1, reproduced below, illustrates a voice-content control device according to one embodiment. *Id.* at 2:36–37, 2:62–64.

FIG.1

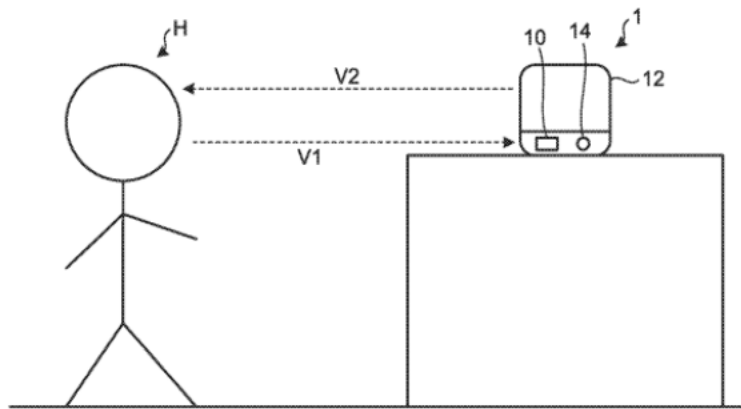


Figure 1 illustrates voice-content control device 1 that detects voice V1 spoken by user H using voice detecting unit 10, analyzes detected voice V1 to perform pre-determined processing, and outputs voice V2 using voice output unit 12. *Id.* at 2:64–3:1. More specifically, voice-content control device 1 analyzes voice V1 and adjusts text to output as voice V2, “thereby suppressing influence of the voice V2 on people other than the user H and

allowing the user H to adequately understand content of the voice V2.” *Id.* at 3:7-11.

Figure 2, reproduced below, illustrates a schematic block diagram of the voice-content control device according to one embodiment. Ex. 1001, 2:38–39, 3:12–13.

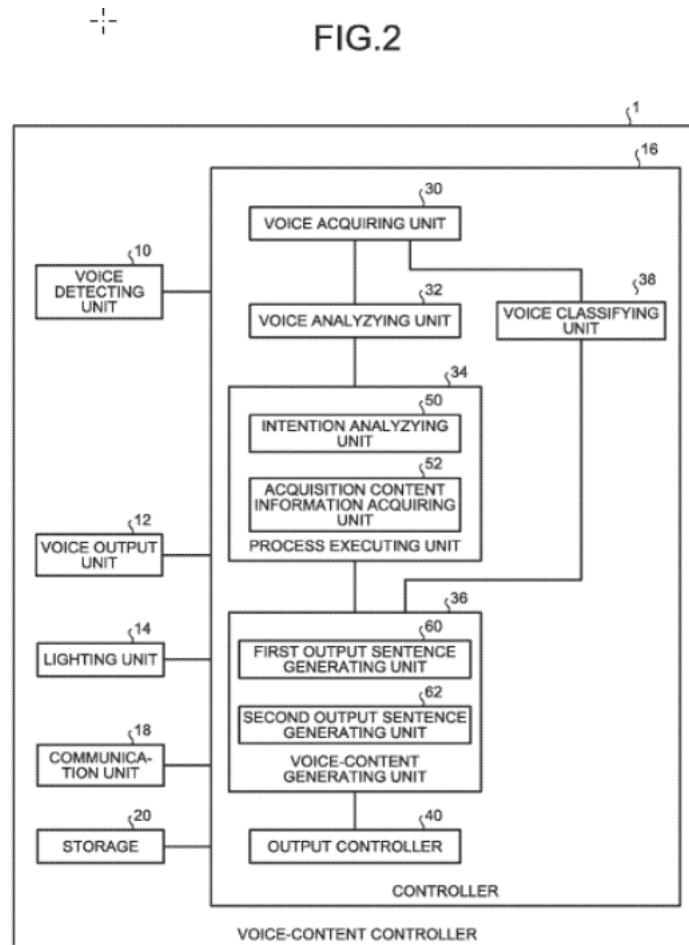


Figure 2 illustrates voice-content control device 1 that includes voice detecting unit 10, voice output unit 12, lighting unit 14, controller 16, communication unit 18, and storage 20. *Id.* at 3:13–7. Controller 16 operates as a central processor unit (“CPU”) because it includes, among other things, voice acquiring unit 30, voice analyzing unit 32, process

executing unit 34, voice-content generating unit 36, voice classifying unit 38, and output controller 40. *Id.* at 3:50–59.

D. Challenged Claims

Of the challenged claims, claims 1, 4, and 5 are independent. Independent claim 1 is directed to “[a] voice-content control device,” independent claim 4 is directed to “[a] voice-content control method,” and independent claim 5 is directed to “[a] non-transitory storage medium that stores a voice-content control program that causes a computer to execute” certain steps. Ex. 1001, 18:39–19:3, 19:22–20:9, 20:10–41. Claims 2 and 3 directly depend from independent claim 1. *Id.* at 19:4–21. Independent claim 1 is illustrative of the challenged claims and is reproduced below.

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

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

1[g] generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice, wherein

1[h] the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto.

Id. at 18:39–19:3 (with bracket annotations provided by Amazon in the Petition).

E. Asserted Prior Art References

Amazon relies on the prior art references set forth in the tables below.

Name¹	Reference	Dates	Exhibit No.
Shin	US 2017/0083281 A1	published Mar. 23, 2017; filed Sept. 19, 2016	1003
Shimomura (with English translation)	JP 2005-202076 A1	published Oct. 28, 2005; filed Jan. 14, 2004	1004
Kristjansson	US 10,147,439 B1	issued Dec. 4, 2018; filed Mar. 30, 2017	1013

F. Asserted Grounds of Unpatentability

Amazon challenges claims 1–5 of the '337 patent based on the asserted grounds of unpatentability set forth in the table below. Pet. 6, 8–68.

¹ For clarity and ease of reference, we only list the first named inventor.

Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
1–5	102(a)(1)/102(a)(2) ²	Shin
1–5	103	Shin
1–5	103	Shimomura
1–5	103	Shimomura, Shin
1–5	103	Shin, Shimomura, Kristjansson

II. ANALYSIS

A. Claim Construction

In this IPR, claim terms are construed using the same claim construction standard as in a civil action under 35 U.S.C. § 282(b). *See* 37 C.F.R. § 42.100(b). That is, claim terms generally are construed in accordance with their ordinary and customary meaning, as would have been understood by a person of ordinary skill in the art, in view of the specification and the prosecution history pertaining to the patent at issue. *See id.* The ordinary and customary meaning of a claim term “is its meaning to the ordinary artisan after reading the entire patent,” and “as of the effective filing date of the patent application.” *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313, 1321 (Fed. Cir. 2005) (en banc).

In its Petition, Amazon contends that “[n]o claim terms require construction to resolve the [unpatentability] challenges [presented] here.” Pet. 8 (citations omitted). Even though Amazon took the position that no

² The Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112-29, 125 Stat. 284, 287–88 (2011), amended 35 U.S.C. §§ 102 and 103, effective March 16, 2013. Because the application that issued as the ’337 patent was filed after this date, the AIA version of §§ 102 and 103 applies. Ex. 1001, codes (21), (22).

claim terms require construction, the Board panel determined in the Institution Decision that the arguments presented by SoundClear in its Preliminary Response warranted preliminarily construing the claim term “a user” as meaning “one or more users.” Dec. on Inst. 14–15 (citing Prelim. Resp. 23–24); *see also id.* at 16 (emphasizing the Board panel’s preliminary construction only requires “a single user” or “at least one user”). Likewise, the Board panel preliminarily construed the claim phrase “a voice spoken by a user,” as recited in limitation 1[b], as covering “one or more voices spoken by a single user, or by plural users.” *Id.* at 15. The Board panel then invited the parties “to fully develop this claim construction issue during trial.” *Id.* at 16.

In its Patent Owner Response, SoundClear recognizes that the Board panel in the Institution Decision preliminarily construed the claim term “a user.” PO Resp. 5 (“[I]n the Institution Decision . . . , the Board *sua sponte* stated that claim construction of a ‘user’ was necessary.” (citing Dec. on Inst. 14)). SoundClear, however, argues that further developing the construction of this claim term is “unnecessary” because “[SoundClear] does not contest the Board’s conclusion that the claims recited ‘one or more users.’” *Id.* at 5–6. Stated differently, SoundClear argues that “no construction of ‘user’ is necessary” because “neither party contests that the scope of the claims extend to a single user speaking in multiple voices.” *Id.* at 6.

Next, SoundClear presents arguments that the Board panel’s preliminary construction of the claim term “a user” impacts the scope and meaning of other claim terms or phrases as follows: (1) it would drastically broaden the claims by converting the claim term “classify” into “a mere

measurement” as opposed to a “classification, or distinction between, at least two separately identifiable voices” (PO Resp. 6 (citing Ex. 2013 ¶¶ 42, 43)); (2) it would render the claim term “voice” superfluous by reducing it to a simple metric like “volume” or “distance” (*id.* at 7–17 (citing Ex. 2013 ¶¶ 44–55; Ex. 2015, 28:12–18, 31:11–32:6, 54:1–6, 59:13–50, 60:3–62:16)); and (3) the claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device” includes three separate and distinct requirements, and it would be improper to read this claim phrase as only requiring classifying the acquired voice as either a first voice or second voice based on distance between the user and the device, irrespective of performing any analysis on the acquired voice (*id.* at 17–21 (citing Ex. 1001, claims 1, 4, 5; Ex. 2013 ¶¶ 56–59; Ex. 2015, 36:7–37:6)).

In its Reply, Amazon contends that we should maintain the Board panel’s preliminary construction of the claim term “a user” as meaning “one or more users” because it is “necessary to resolve a disputed issue.” Pet. Reply 1; *see also id.* at 6 (“[T]he Board should construe “a user” to mean “one or more users.”). In response to SoundClear’s arguments regarding the impact of this construction on the scope and meaning of other claim terms or phrases, Amazon argues that we need not construe the claim term “classify” because the plain language of the claims and the specification both support performing classification based on distance alone. *Id.* at 3–4 (citing Ex. 1001, 8:10–12, 8:20–26, 8:30–34, claim 1; Ex. 1202 ¶¶ 6, 7). Amazon further argues that SoundClear’s arguments regarding the claim term “volume” are “irrelevant” because the unpatentability arguments presented in the Petition do not “depend on volume classification.” *Id.* at 5. Lastly, Amazon argues that SoundClear’s arguments regarding the claim

phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device” is an attempt to rewrite the claims. *Id.* at 7. Amazon admits that “the distance calculation is not the claimed analysis” because “[t]he proximity sensor calculates distance.” *Id.* Instead, Amazon asserts that the analysis performed by the claimed “voice classifying unit” “includes comparing the voices’ distance to a threshold, and the result is classification as a ‘first voice’ or ‘second voice.’” *Id.* (citing Ex. 1001, 8:20–34; Ex. 1202 ¶ 15).

In its Sur-reply, SoundClear contends that “the patentee chose the clearest, most straightforward language available to convey that the voice is the object that must be classified: ‘classify the voice.’” PO Sur-reply 3 (footnote omitted). SoundClear argues that the parties agree that the claim term “voice” is distinct from the claim terms “volume” and “distance.” *Id.* at 3–4 (citing Ex. 2013 ¶¶ 44–55; Ex. 2015, 28:12–18, 31:11–32:6, 54:1–6. Given this distinction, SoundClear asserts that “[t]he plain meaning of ‘classify the voice’ cannot be expanded to ‘classify the distance’” because “it would lead to implausible results.” *Id.* at 5.

Based on the complete record, we maintain the Board panel’s preliminary construction of the claim term “a user” as meaning “one or more users.” Dec. on Inst. 15. With this construction in mind, there appears to be no dispute between the parties that the scope and meaning of this claim term extends to a single user speaking in multiple voices. *See* PO Resp. 6. Instead, the dispute between the parties turns on how the construction of the claim term “a user” impacts the scope and meaning of other claim terms or phrases, including the claim terms “classify” and “voice,” as well as how these terms impact the claim phrase “analyze a voice . . . to classify the

voice . . . based on distance between the user and the . . . device.” *Compare* PO Resp. 6–21, *with* Pet. Reply 1–9.

To be clear, contrary to SoundClear’s claim construction arguments, the Board panel in the Institution Decision did not preliminarily construe the claim terms “classify” and “voice,” much less the claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device.” *See* Dec. on Inst. 14–16. For the first time in the Patent Owner Response, SoundClear raises arguments that the Board panel’s preliminary construction of the claim term “a user,” which SoundClear does not dispute, impacts the scope and meaning of these other claim terms and phrase. *See* PO Resp. 6–21. But there is nothing ambiguous or linguistically obscure about the scope and meaning of the aforementioned claim terms and phrase that would warrant deviating from their ordinary and customary meanings. *See Senmed, Inc. v. Richard–Allan Med. Indus., Inc.*, 888 F.2d 815, 819 n.8 (Fed. Cir. 1989) (“Lawyers may create a ‘dispute’ about any word, but there is nothing ambiguous or linguistically obscure about [the claim terms and phrases] as used in the present claim.”).

As an example, the claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device” is recited in independent claim 1, and similarly recited independent claims 4 and 5. Ex. 1001, 18:42–46 (claim 1), 19:26–29 (claim 4), 20:20:16–19 (claim 5). In our view, the plain language of this claim phrase controls because it merely requires analyzing a voice spoken by a user that is acquired by a device to classify the voice as either a first or second voice based on the distance between the user and the device. *See Bell Commc’ns Research, Inc. v. Vitalink Commc’ns Corp.*, 55 F.3d 615, 619–620 (Fed. Cir. 1995)

(explaining that we must first look to the words of the claims, which should be accorded their ordinary and customary meaning unless it appears the patentees used them otherwise); *see also* Tr. 25:22–25:8 (upon inquiring whether the intrinsic evidence, such as the plain language of the challenged claims, was controlling, counsel for SoundClear stated “you don’t even have to go outside the claims because it’s in the claim. I mean, it’s recited in the claim, analyze a voice.”), 40:14–19 (counsel for SoundClear agreeing that intrinsic evidence is controlling).

The specification of the ’337 patent further reinforces that the plain language of the claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device” is controlling. *See United States v. Adams*, 383 U.S. 39, 49 (1966) (stating that it is equally “fundamental that claims are to be construed in the light of the specification and both are to be read with a view to ascertaining the invention” (citations omitted)). The specification discloses that “voice classifying unit 38 analyzes the voice V1 acquired by the voice acquiring unit 31, and classifies the voice V1 to either one of a first voice V1A and a second voice V1B.” Ex. 1001, 7:51–54. The specification makes clear that voice classifying unit 38 performs the functions of analyzing the voice acquired by voice acquiring unit 31 to classify the voice as either a first voice or a second voice “by using any method.” *Id.* at 8:10–12. One such method includes using a proximity sensor to measure a distance between user H and voice content-control device 1 illustrated in Figure 1, and then using the distance “as a feature value to perform the classification.” *Id.* at 8:20–26. In the particular case where distance is used as the feature value, “voice classifying unit 38 sets a threshold of the feature value [(i.e., distance)], and classifies the voice

V1 as either one of the first voice V1A or the second voice V1B based on whether the feature value [(i.e., distance)] exceeds the threshold.” *Id.* at 8:30–34.

Indeed, the patentee of the ’337 patent had an opportunity during prosecution to further limit the challenged claims by requiring them to perform the claimed analysis using distance in conjunction with other feature values, but declined to do so. *See* Ex. 2009, 154–156. Instead, upon entering a request for continued examination, the patentee amended each of the claims that issued as independent claims 1, 4, and 5 of the ’337 patent by requiring that the claimed analysis be “based on the distance between the user and the voice-content control device.” *Id.* (amending independent claims 1, 5, and 6 that issued as independent claims 1, 4, and 5 of the ’337 patent).

Contrary to SoundClear’s claim construction arguments, we need not distill the claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device” into the following requirements: (1) a classification that must be based on (2) an analysis of the acquired voice and (3) a distance, which is calculated by a proximity sensor. *See* PO Resp. 17–21. This would rewrite this claim phrase to include three separate and distinct requirements. *See K-2 Corp. v. Salomon S.A.*, 191 F.3d 1356, 1364 (Fed. Cir. 1999) (“Courts do not rewrite claims; instead, [they] give effect to the terms chosen by the patentee.”). Moreover, reading this claim phrase to require three separate and distinct requirements either ignores or overlooks the plain language of the claims, the specification, and the prosecution history, each of which, as we explain above, support performing the functions of analyzing a voice spoken by a

user that is acquired by a device to classify the voice as either a first voice or a second voice based on the distance between the user and the device (e.g., by using a distance threshold calculation).

In summary, we maintain the Board panel’s preliminary construction of the claim term “a user” as meaning “one or more users” because there is no dispute between the parties that the scope and meaning of this claim term extends to a single user speaking in multiple voices. We, however, decline to construe separately the claim terms “classify” and “voice,” as well as how these terms impact the claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device.” Instead, we construe the aforementioned claim terms and phrase in accordance with their ordinary and customary meanings.

B. Anticipation by and Obviousness over Shin

Amazon contends that claims 1–5 of the ’337 patent are unpatentable under (1) 35 U.S.C. §§ 102(a)(1)/102(a)(2) as anticipated by Shin or (2) 35 U.S.C. § 103 as obvious over the teachings of Shin. Pet. 8–35. Amazon contends that Shin discloses the subject matter of each challenged claim and, only with respect to the obviousness ground based on Shin, Amazon provides reasoning as to why a person of ordinary skill in the art would have been prompted to modify the teachings of Shin. *Id.* To support its positions, Amazon submits a Declaration of Richard Stern, Ph.D., with the Petition and a Reply Declaration of Dr. Stern with the Reply. See Exs. 1002, 1202.

SoundClear contends that Shin does not disclose limitations 1[b], 1[f], 1[g], and 1[h]. PO Resp. 23–33; PO Sur-reply 5–15. In addition, with respect to the anticipation ground based on Shin, SoundClear argues that

Shin does not disclose all the components as arranged in independent claim 1. PO Resp. 33–35. With respect to the obviousness ground based on Shin, SoundClear argues that Amazon does not provide a sufficient rationale to modify the teachings of Shin. PO Resp. 35–37; PO Sur-reply 18–19. To support its positions, SoundClear submits the Declaration of Joseph McAlexander III, P.E., with the Patent Owner Response. *See* Ex. 2013.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1–5 of the ’337 patent are unpatentable as anticipated by Shin. We begin our analysis with the principles of law that generally apply to an asserted ground based on anticipation and obviousness, we then assess the level of skill in the art, next we provide an overview of Shin, and then we turn to the parties’ contentions with respect to the challenged claims.

1. Principles of Law

To establish anticipation, “all of the elements and limitations of the claim must be shown in a single prior reference, arranged as in the claim.” *Karsten Mfg. Corp. v. Cleveland Golf Co.*, 242 F.3d 1376, 1383 (Fed. Cir. 2001). When evaluating a single prior art reference in the context of anticipation, the reference must be “considered together with the knowledge of one of ordinary skill in the pertinent art.” *In re Paulsen*, 30 F.3d 1475, 1480 (Fed. Cir. 1994) (citing *In re Samour*, 571 F.2d 559, 562 (CCPA 1978)). “[T]he dispositive question regarding anticipation [i]s whether one skilled in the art would reasonably understand or infer from the [prior art reference’s] teaching’ that every claim element was disclosed in that single reference.” *Dayco Prods., Inc. v. Total Containment, Inc.*, 329 F.3d 1358, 1368 (Fed. Cir. 2003) (alterations in original) (quoting *In re Baxter Travenol*

Labs., 952 F.2d 388, 390 (Fed. Cir. 1991)). We analyze the asserted ground based on anticipation with the principles we identify above in mind.

A claim is unpatentable as obvious “if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains.” 35 U.S.C. § 103; *see KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007). The question of obviousness is resolved on the basis of underlying factual determinations, including (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) when in evidence, objective indicia of obviousness or non-obviousness (i.e., secondary considerations, such as commercial success, long-felt but unsolved need, failure of others, etc.). *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1, 17–18 (1966). We analyze the asserted grounds based on obviousness with the principles we identify above in mind.

2. *Level of Skill in the Art*

In determining whether an invention would have been obvious, we consider the level of ordinary skill in the pertinent art at the relevant time. *Graham*, 383 U.S. at 17. “The importance of resolving the level of ordinary skill in the art lies in the necessity of maintaining objectivity in the obviousness inquiry.” *Ryko Mfg. Co. v. Nu-Star, Inc.*, 950 F.2d 714, 718 (Fed. Cir. 1991). The “person having ordinary skill in the art” is a hypothetical construct, from whose vantage point obviousness is assessed. *In re Rouffet*, 149 F.3d 1350, 1357 (Fed. Cir. 1998).

Factors pertinent to a determination of the level of ordinary skill in the art include “(1) the educational level of the inventor; (2) type of problems encountered in the art; (3) prior art solutions to those problems; (4) rapidity with which innovations are made; (5) sophistication of the technology; and (6) educational level of active workers in the field.” *Env’t Designs, Ltd. v. Union Oil Co. of Cal.*, 713 F.2d 693, 696 (Fed. Cir. 1983) (citing *Orthopedic Equip. Co. v. All Orthopedic Appliances, Inc.*, 707 F.2d 1376, 1381–82 (Fed. Cir. 1983)). “Not all such factors may be present in every case, and one or more of these or other factors may predominate in a particular case.” *Id.* at 696–97.

Relying on the testimony of Dr. Stern, Amazon argues the following:

a person of ordinary skill in the art . . . would have had a minimum of a bachelor’s degree in computer engineering, computer science, electrical engineering, or a similar field, and approximately two years of industry or academic experience in a field related to controlling the audio output of electronic devices. Work experience can substitute for formal education and additional formal education can substitute for work experience.

Pet. 7 (citing Ex. 1002 ¶¶ 30–34). Relying on the testimony of Mr. McAlexander, SoundClear argues the following:

[a person of ordinary skill in the art] at the time of the invention of the ’337 patent would have . . . a Bachelor of Science degree in electrical or computer engineering (or a related academic field), and at least two (2) additional years of experience in the design and development of electronic devices or systems related to controlling the audio output based on inputs.

PO Resp. 2 (citing Ex. 2013 ¶ 32).

We do not discern a material difference between the assessment of the level of skill in the art advanced by each party, nor does either party premise

its arguments exclusively on its own assessment. Based on the complete record, we adopt Amazon's assessment of the level of skill in the art because (1) it is consistent with the '337 patent and the asserted prior art, and (2) it is supported by the testimony of Dr. Stern. We note, however, that both our anticipation and obviousness analysis would be the same under SoundClear's assessment.

3. Overview of Shin

Shin generally relates to “voice input for an electronic device” and, in particular, “to a method and an electronic device for providing content in response to a voice input.” Ex. 1003 ¶ 3. Shin discloses performance drawbacks associated with providing content based on electrical signals converted from a user's speech received as voice input, such as lack of an emotional connection between the user and the electronic device when providing content in response to the voice input. *Id.* ¶¶ 6, 7. For example, Shin discloses that “the electronic device provides the content with a uniform speed, a monotonous tone, and a preset volume, regardless of the user's conditions while providing the voice input,” and because “the electronic device does not consider nuances according to a form of the user's speech, it is difficult for the electronic device to provide content appropriate for the user's condition.” *Id.* ¶ 7. To address these performance drawbacks, Shin discloses “a content providing method that analyzes a voice input of a user's speech and determine[s] an output scheme of content based on the various speech features obtained from results of the analysis.” *Id.* ¶ 8.

Figure 3 of Shin, reproduced below, illustrates an electronic device (e.g., smartphones, tablet personal computers (“PCs”), mobile phones, etc.)

in accordance with one embodiment. Ex. 1003 ¶¶ 15, 45; *see also id.* ¶ 31 (disclosing examples of what may constitute an electronic device).

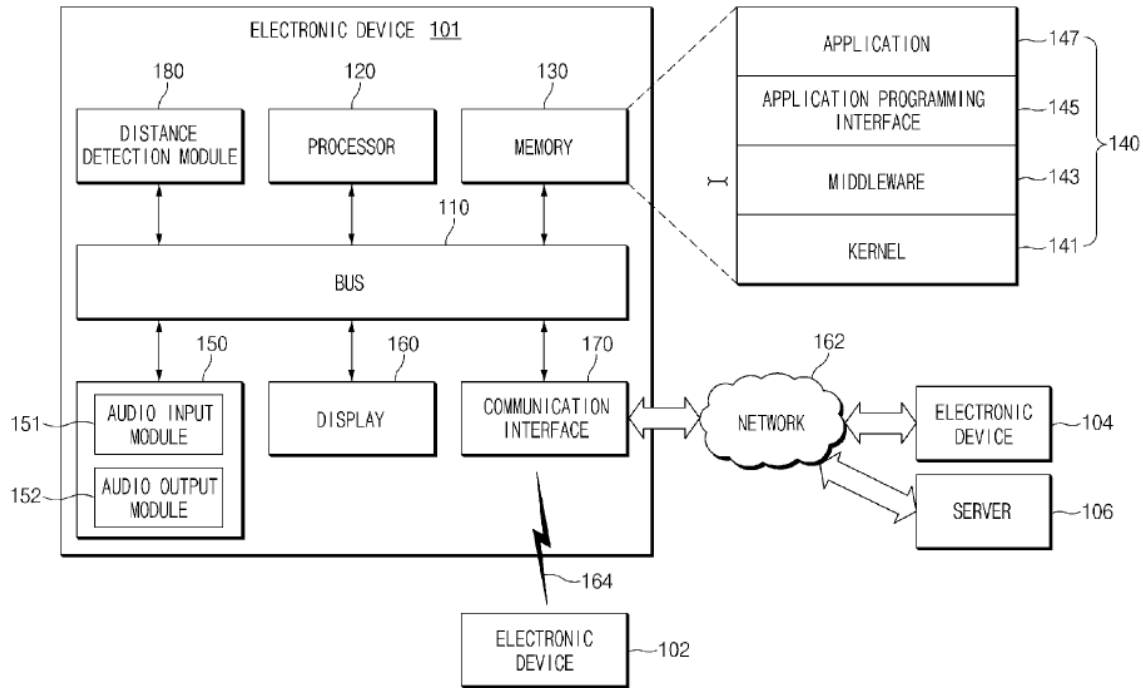


FIG.3

Figure 3 illustrates electronic device 101 that includes, among other things, processor 120 that may execute a voice recognition application to analyze voice input received by audio input module 151, distance detection module 180 that may include a sensor to calculate a distance between a user and electronic device 101, and audio output module 151 that outputs content corresponding to the voice input. *Id.* ¶¶ 45, 47, 48, 66. Shin’s processor 120 is capable of tailoring output content based on the distance between a user and electronic device 101 itself. *Id.* ¶¶ 51, 53. As one example, processor 120 may output certain content for a user within 1 meter (“m”) of electronic device 101 and output abbreviated content for a user more than 1 m away from electronic device 101. *Id.* ¶¶ 86–88, Table 3.

4. Claims 1, 4, and 5

Our analysis below focuses on the limitations of independent claim 1 by relying on the bracketed annotations provided by Amazon in the Petition. *See* Pet. i, 8–24. For the most part, the same arguments and evidence Amazon provides to explain how Shin discloses all the limitations of independent claim 1 equally apply to the corresponding limitations of independent claims 4 and 5. *See id.* at 29–30.

Limitation 1[pre] recites “[a] voice-content control device comprising.” Ex. 1001, 18:39. To the extent the preamble is treated as limiting, Amazon contends that Shin discloses the features recited in the preamble because it describes an electronic device (e.g., electronic device 101 illustrated in Figure 3), which may be a smartphone, PC, or home appliance, that receives a user’s voice input through an embedded audio input module (e.g., a microphone), “generate[s] content corresponding to a result of analyzing the voice input,” and “provide[s] the generated content as a sound through an embedded audio output module (e.g., a speaker).” Pet. 8–9 (quoting Ex. 1003 ¶ 37) (citing Ex. 1003, Abstract, ¶¶ 8–11, 31, 38, 39, 48–60, 69–72, claims 1, 15, 20, Fig. 1; Ex. 1002 ¶ 51). As one example, if a user says “Let me know what time it is now,” Shin’s electronic device responds with “The current time is nine ten AM.” *Id.* at 8 (quoting Ex. 1003 ¶ 37).

Limitation 1[a] recites “a proximity sensor configured to calculate a distance between a user and the voice-content control device.” Ex. 1001, 18:40–41. Amazon contends that Shin discloses this limitation because its electronic device includes a “distance detection module,” that “compute[s] a distance between a user and the electronic device.” Pet. 9 (quoting Ex. 1003

¶ 66) (citing Ex. 1003 ¶¶ 53, 75, 77, Fig. 5A; Ex. 1002 ¶ 53). Amazon argues that Shin’s distance detection module may include “various sensors,” such as “an infra-red sensor, [a radio frequency] sensor, an ultrasonic sensor, and the like.” *Id.* at 9–10 (quoting Ex. 1003 ¶ 66) (citing Ex. 1003 ¶ 42; Ex. 1002 ¶ 54). As an example of an electronic device that includes a distance detection module, Amazon points to Shin’s Figure 3 as illustrating electronic device 101 that includes, among other things, distance detection module 180. *Id.* at 10 (citing Ex. 1003 ¶¶ 45, 53, 54, 77, 96, Fig. 3, claims 5, 9; Ex. 1002 ¶¶ 55, 56). As another example of an electronic device that includes a proximity sensor, Amazon points to Shin’s Figure 9 as illustrating electronic device 901 that includes, among other things, proximity sensor 940G. *Id.* (citing Ex. 1003 ¶¶ 143, 153, Fig. 9; Ex. 1002 ¶ 57).

Limitation 1[b] recites “a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device.” Ex. 1001, 18:42–46. Amazon contends that Shin’s Figure 3 illustrates electronic device 101 that includes, among other things, processor 120 and audio input module 151. Pet. 12 (citing Ex. 1003 ¶ 45, Fig. 3; Ex. 1002 ¶ 62). Amazon argues that Shin’s processor 120 (e.g., a CPU) discloses the claimed “voice classifying unit” because it “analyze[s] a voice input through [the] audio input module 151.” *Id.* (quoting Ex. 1003 ¶ 48) (citing Ex. 1003, Abstract, ¶¶ 3–11, 29, 37–39, 41, 60, 61, claims 1, 15, 20). Amazon argues that Shin’s audio input module 151, which may be “implemented with a microphone or the like,” discloses the claimed “voice acquiring unit”

because it “obtain[s] a user’s speech as a voice input.” *Id.* (quoting Ex. 1003 ¶ 62) (citing Ex. 1003 ¶¶ 37, 155, 158).

Amazon further contends that Shin’s processor 120 “analyzes the voice to classify it as either a first or second voice based on the user-to-device distance” because it executes a voice recognition application to process voice input, and then it generates a corresponding “output scheme,” which includes the content and volume level, “based on the distance between the user and the electronic device 101.” Pet. 13 (quoting Ex. 1003 ¶¶ 78, 86) (citing Ex. 1003 ¶¶ 37, 39, 47, 53, 60, 77–80, claims 6, 17; Ex. 1002 ¶ 63)). Amazon argues that an example of Shin’s processor 120 analyzing the voice to classify it based on the user-to-device distance is provided in a consolidated, annotated version of Shin’s Table 3, reproduced below.

TABLE 3		
	Distance (D) between user and electronic device	Output amount of information of content
First Voice	$D < 1\text{ m}$	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
Second Voice	$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

Id. at 13–14 (citing Ex. 1003, Table 3). Amazon argues that its consolidated, annotated version of Shin’s Table 3 illustrates “classifying the voice as a ‘first voice’ when the distance is less than 1 meter, and a ‘second voice’ when the distance is greater than 4 meters” (both in red). *Id.* at 14 (citing Ex. 1003 ¶¶ 51, 79–81, 86, 87, Table 1; Ex. 1002 ¶ 64).

Alternatively, Amazon asserts that “[a]ny distance range up to 4 meters would satisfy the claimed ‘first voice.’” *Id.* (citing Ex. 1002 ¶ 65).

Limitation 1[c] recites “a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user.” Ex. 1001, 18:47–49. Amazon contends that Shin’s processor 120 discloses the claimed “process executing unit” because it analyzes the voice input acquired from audio input module 151, it executes a voice recognition application to process the voice input (which requests content of interest to the user), and it generates corresponding content for output. Pet. 15 (citing Ex. 1003 ¶¶ 37, 48, 58–60, 62, 67, 71, 84, 88, 89, 103, 107, 120, 126, 127, 141; Ex. 1002 ¶ 68). For example, Amazon argues that Shin’s processor 120 analyzes the acquired voice input “to execute processing required by the user” in the same way as the ’337 patent because Shin discloses analyzing a user’s speech (“Let me know today’s weather”) to execute processing to obtain the required weather information. *Id.* at 16 (citing Ex. 1003 ¶ 88; Ex. 1002 ¶ 69).

Limitation 1[d] recites “a voice-content generating unit configured to generate, based on content of the processing executed by the processing executing unit, output sentence that is text data for a voice to be output to the user.” Ex. 1001, 18:50–53. Relying on the same arguments and evidence it presents for limitations 1[b] and 1[c], Amazon contends that Shin’s processor 120 discloses the claimed “voice-content generating unit” because it generates the content for output to the user. Pet. 17 (citing Ex. 1002 ¶ 72). For example, Amazon argues that content generated by Shin’s processor 120 can be a sentence including text data, such as “The current time is nine ten AM” or “The weather in Jul. 1 2025 is rainy after cloudy.” *Id.* (first quoting

Ex. 1003 ¶ 37, then quoting Ex. 1003 ¶ 88). Amazon asserts that Shin converts this content into a voice using text-to-speech technology and outputs it audibly. *Id.* (citing Ex. 1003 ¶¶ 7, 37, 39, 48, 62, 72, 103, 131, 139, 158; Ex. 1002 ¶¶ 71, 72). Amazon also argues that Shin’s processor 120 discloses the claimed generating the output sentence “based on content of the processing executed by the processing executing unit” because it analyzes the voice input to generate content that aligns with the user’s interest. *Id.* at 17–18.

Limitation 1[e] recites “an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto.” Ex. 1001, 18:54–56. Relying on the same arguments and evidence it presents for limitation 1[d], Amazon contends that Shin discloses this limitation because processor 120 converts content to voice, and then “adjust[s]” the “output volume level.” Pet. 18 (quoting Ex. 1003 ¶ 53) (citing Ex. 1003 ¶¶ 76–81, 98–101, 105, claims 6, 17).

Limitation 1[f] recites “the voice-content generating unit is further configured to generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice.” Ex. 1001, 18:57–60. Amazon contends that Shin discloses this limitation because processor 120 adjusts the amount of information provided to the user based on the user-to-device distance. Pet. 19 (citing Ex. 1003 ¶¶ 51, 53, 86–90, claims 8, 17; Ex. 1002 ¶ 77). Amazon argues that one example of Shin’s processor 120 generating a first output sentence when the acquired voice has been classified as the first voice is provided in a consolidated, annotated version of Table 3, reproduced below.

TABLE 3

	Distance (D) between user and electronic device	Output amount of information of content
First Voice	D < 1 m	Amount of information of original content
	1 m ≤ D < 2 m	Content of first abbreviation level
	2 m ≤ D < 3 m	Content of second abbreviation level
	3 m ≤ D < 4 m	Content of third abbreviation level
	D ≥ 4 m	Content of fourth abbreviation level (abstract content)

Id. (citing Ex. 1003 ¶¶ 51, 87, Table 3). Amazon argues that its consolidated, annotated version of Shin’s Table 3 illustrates that, “when classified as a first voice (e.g., closer than 1 meter [to electronic device 101]), the processor [120] outputs the content without abbreviation” (in red). *Id.* (citing Ex. 1003 ¶ 87; Ex. 1002 ¶ 77). In the weather example identified previously, Amazon asserts that, when Shin’s processor 120 classifies the voice as a first voice (e.g., less than 1 meter to electronic device 101), the output content is the “original content,” which includes outputting “The weather in Jul. 1, 2015 is [(1)] rainy after cloudy [(2)]. The highest temperature is 28° [Celsius (‘C’)]., and the lowest temperature is 18° C., [(3)], the rainfall is 10 [millimeters] [(4)]. Prepare your umbrella when you go out. [(5)].” *Id.* at 19–20 (quoting Ex. 1003 ¶ 88).

Limitation 1[g] recites the voice-content generating unit is further configured to “generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as a second voice.” Ex. 1001, 18:61–64. Amazon contends that Shin discloses this limitation because, whenever the user-to-device “distance increases by 1 m, electronic device 101 may decrease the output amount of information by making the abbreviation level

of the original content high.” Pet. 20 (quoting Ex. 1003 ¶ 87) (citing Ex. 1003 ¶ 86). Amazon argues that one example of Shin’s processor 120 generating a second output sentence when the acquired voice has been classified as the second voice is provided in a consolidated, annotated version of Table 3, reproduced below.

TABLE 3		
	Distance (D) between user and electronic device	Output amount of information of content
First Voice	$D < 1\text{ m}$	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
Second Voice	$D \geq 4\text{ m}$	Content of fourth abbreviation level (abstract content)

Id. at 20–21 (citing Ex. 1003, Table 3; Ex. 1002 ¶ 81). Amazon argues that its consolidated, annotated version of Shin’s Table 3 illustrates that, when the user’s voice is classified as a second voice (e.g., more than 4 meters from electronic device 101), the voice falls into a second classification and processor 120 outputs “[c]ontent of fourth abbreviation level (abstract content)” (in red). *Id.* at 21 (quoting Ex. 1003 ¶ 86³) (citing Ex. 1003 ¶ 51; Ex. 1002 ¶ 81). In the weather example identified previously, Amazon asserts that, when Shin’s processor 120 classifies the voice as a second voice (e.g., more than 4 meters from electronic device 101), the output content omits some information, which includes abbreviating the original content by

³ Amazon inadvertently cited to paragraph 87 of Shin, but the quoted text actually appears in Table 3 itself, which is reproduced in paragraph 86 of Shin.

simply outputting “rainy after cloudy.” *Id.* at 21–22 (citing Ex. 1003 ¶ 88; Ex. 1002 ¶ 82).

Limitation 1[h] recites “the output controller is further configured to adjust the sound volume of the voice data such that the sound volume of the voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto.” Ex. 1001, 18:65–19:3. Relying on the same arguments and evidence it presents for limitation 1[e], Amazon contends that Shin discloses this limitation because processor 120 adjusts the content’s output volume based on the user-to-device distance. Pet. 22 (citing Ex. 1003 ¶¶ 53, 77–81). Amazon argues that one example of Shin’s processor 120 adjusting the sound volume of voice data is provided in Table 1, reproduced below.

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

Id. at 23 (citing Ex. 1003, Table 1). Amazon argues that Shin’s Table 1 illustrates that “electronic device 101 may set an output volume level of the content to 40 [decibel (“dB”)] when the distance between the user and the electronic device 101 is less than 1 m and may increase the output volume level by 5 dB whenever the distance increases by 1 m.” *Id.* at 22–23 (quoting Ex. 1003 ¶ 80). According to Amazon, this means that, when the distance between the user and electronic device 101 is less than 1 meter, the output volume level is 40 dB, and when the distance between the user and

electronic device is more than 4 meters, the output volume level is 60 or 65 dB. *Id.* at 23 (citing Ex. 1003 ¶¶ 49, 53, 88, 140; Ex. 1002 ¶¶ 87, 88).

For purposes of the asserted ground based on obviousness over Shin, Amazon contends that, even if SoundClear were to argue that independent claim 1 “requires the user’s voice to be classified into no more than two classifications based on distance,” Shin would still render the subject matter of independent claim 1 obvious because a person of ordinary skill in the art would have implemented Shin’s Table 1 (i.e., the volume change table) and Shin’s Table 3 (i.e., the output information table) with only two distance ranges. Pet. 30 (citing Ex. 1002 ¶¶ 108–116).

Amazon provides six reasons as to why a person of ordinary skill in the art would have been motivated to modify Shin to include no more than two classifications based on distance. Pet. 30–32. As one example, Amazon argues that Shin explicitly discloses dividing content into two classifications so a person of ordinary skill in the art would have been motivated to “use two volume levels, one for each classification.” *Id.* at 30 (citing Ex. 1003 ¶ 51; Ex. 1002 ¶ 109). Amazon asserts that a person of ordinary skill in the art would have had a reasonable expectation of success in modifying Shin in this way because “[d]oing so would have been trivial, involving configuring Shin’s processor [120] with revised tables for two proximity-based categories and optimizing performance for specific users.” *Id.* (citing Ex. 1002 ¶ 115).

Notwithstanding SoundClear’s arguments to the contrary, Amazon has shown by a preponderance of the evidence that Shin discloses all the limitations of independent claims 1, 4, and 5. During trial, SoundClear contends that Shin does not disclose limitations 1[b], 1[f], 1[g], and 1[h].

PO Resp. 23–33; PO Sur-reply 5–15. In addition, with respect to the anticipation ground based on Shin, SoundClear argues that Shin does not disclose all the components as arranged in independent claim 1. PO Resp. 33–35. With respect to the obviousness ground based on Shin, SoundClear argues that Amazon does not provide a sufficient rationale to modify the teachings of Shin. PO Resp. 35–37; PO Sur-reply 18–19.

We begin our analysis by addressing SoundClear’s arguments directed to limitations 1[b], 1[f], 1[g], and 1[h], next we address SoundClear’s arguments directed to whether Shin discloses the claimed components arranged as required by independent claim 1, and then we address SoundClear’s arguments as to whether Amazon provides a sufficient rationale to modify the teachings of Shin.

a. Limitation 1[b]—“a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device”

In its Patent Owner Response, SoundClear presents the following two arguments as to why Shin does not disclose limitation 1[b]: (1) Shin discloses measurement of distance—not classification of voices (PO Resp. 23–29; PO Sur-reply 5–11); and (2) Shin does not disclose analysis of a voice for classification (PO Resp. 29–32; PO Sur-reply 11–15). Although these two arguments overlap in certain respects, we address them in turn.

i. SoundClear’s First Argument

First, SoundClear contends that Amazon relies on Shin’s measurement of distance between the user and the electronic device to disclose limitation 1[b], but Shin fails to disclose “classifying a voice spoken by a user ‘as either a first voice [and/or] a second voice,’ as claimed.” PO

Resp. 23. According to SoundClear, the only disclosure of “voice” in Shin relates to processing the voice in connection with generating content, but this “garden variety natural language processing to determine user intent” does not amount to classifying a voice as a first or second voice. *Id.* at 24 (citing Ex. 2013 ¶ 68). SoundClear also argues that Shin only mentions “classification” of “content” in the context of output—not voice input. *Id.* SoundClear further argues that Amazon’s reliance on Shin’s Table 3 to demonstrate the claimed “first voice” and “second voice” is misplaced because this table only discloses “a calculation of distance and nothing else.” *Id.* at 24–25 (citing Ex. 2013 ¶¶ 69–71; Pet. 13). Moreover, SoundClear argues that Shin only contemplates a single voice in connection with Table 3 and, therefore, a person of ordinary skill in the art would not have understood that Shin’s calculation of distance between the user and the device to be synonymous with voice or classification. *Id.* at 25–26 (citing Ex. 1002 ¶ 72). Stated differently, SoundClear argues that Shin merely correlates content to distance between the user and the device, but it does not classify the voice output in any way, let alone classify the voice as a first or second voice. *Id.* at 26 (citing Ex. 1002 ¶ 73). Lastly, SoundClear asserts that Amazon’s declarant, Dr. Stern, admits that Shin’s Table 3 does not disclose voice classification, but rather it is limited to information regarding the distance between the user and the electronic device. *Id.* at 27–28 (citing Ex. 2015, 23:5–24:17, 69:19–70:3, 72:12–73:8).

In its Reply, Amazon contends that Shin is replete with disclosures of analyzing voice input to determine an output scheme. Pet. Reply 9 (citing Ex. 1003, Abstract, ¶¶ 48, 39, 53; Pet. 12). As one example, Amazon argues that Shin plainly discloses analyzing voice input because this appears as an

explicit step in the flowchart illustrated in Figure 5A. *Id.* at 9–10 (citing Ex. 1003, Fig. 5A; Ex. 1202 ¶ 17). Amazon further argues that SoundClear’s argument about Shin lacking an explicit disclosure of classification lacks merit because there is no requirement as a matter of law that Shin must use the exact same words as the challenged claims. *Id.* at 10. In addition, Amazon argues that Shin’s Table 3 is not limited to a distance calculation, but rather this table amounts to an explicit disclosure of “*classifications based on distance*” because the left-hand column in Table 3 includes categories or classifications for distance, whereas the right-hand column in Table 3 describes the output resulting from each classification. *Id.* at 10–11 (citing Ex. 1003, Table 3; Pet. 11–14; Ex. 1202 ¶ 18); *see also id.* at 12 (arguing that “Shin’s Table 3 discloses more than just distance determination—it discloses voice classifications based on distance” (citing Ex. 1202 ¶ 18)). Lastly, Amazon argues that SoundClear’s reliance on the deposition testimony of Amazon’s declarant, Dr. Stern, is misleading because he did not admit that Shin’s Table 3 is limited to information regarding the distance between the user and the electronic device. *Id.* at 12–13. According to Amazon, Dr. Stern testified that Shin’s Table 3 discloses voice classification. *Id.* at 13 (citing Ex. 2015, 68:16–69:18, 70:16–71:6, 71:18–72:7).

In its Sur-reply, SoundClear contends that Amazon does not identify how Shin discloses classifying voice input, but rather both Amazon and its declarant, Dr. Stern, improperly rely on unsupported assumptions based on Shin’s inputs and outputs. PO Sur-reply 5–6; *see also id.* at 9 (arguing that Amazon “identif[ies] not express disclosure of any classification of inputs in Shin”). According to SoundClear, Dr. Stern provides deposition testimony

confirming that adjusting an output based on an input does not disclose classifying voice input. *Id.* at 6–8 (citing Ex. 2016, 21:8–22:16, 33:5–34:15, 62:22–63:8, 64:5–65:2, 65:7–12, 77:19–78:3). Based on the cited portions of Dr. Stern’s deposition testimony, SoundClear argues that a person of ordinary skill in the art would not assume that classification necessarily occurs in Shin. *See id.* at 8. At best, SoundClear asserts that Shin discloses classifying distances instead of classifying voices, as claimed. *Id.* at 10–11 (citing Ex. 1202 ¶¶ 14–16, 19, 20, 24, 25).

Based on the complete record, we agree with Amazon that Shin discloses classifying a user’s voice as either a first or second voice based on a distance between the user and the device, as required by limitation 1[b]. Pet. 11–14; Pet. Reply 9–14. In the context of describing the components included within electronic device 101 illustrated in Figure 3, Shin discloses that “audio input module 151 may obtain a user’s speech as a voice input.” Ex. 1003 ¶ 62; *see also id.* ¶ 37 (“For example, the user 10 may speak ‘Let me know what time it is now.’ with a moderate volume and at a moderate speech rate, and the speech may be provided . . . as voice input.”). Shin’s “processor 120 may analyze [the] voice input received through . . . audio input module 151.” *Id.* ¶ 48. According to Shin, processor 120 may include one or more CPUs, each of which “may execute voice recognition application . . . to perform the content providing method.” *Id.* ¶ 47; *see also id.* ¶ 60 (disclosing that “memory 130 may store software and/or a program 140,” which includes, among other things, application program 147 that further “include[s] voice recognition application . . . that is capable of analyzing voice input, and generating and/or providing the content corresponding to the voice input”).

In the context of describing the flowchart representing the content providing method illustrated in Figure 5A, Shin discloses that processor 120 “determines an output scheme of content based on the distance between the user and the electronic device 101.” Ex. 1003 ¶ 78, Fig. 5A (step 505). One example of the output scheme provided by processor 120 is illustrated in a consolidated view of Shin’s Table 3, reproduced below.

TABLE 3

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

Id. ¶ 86. This consolidated view of Table 3 illustrates that, “when a distance between the user and the electronic device 101 is less than 1 m, an output amount of information of the content may be set highest.” *Id.* ¶ 87. That is, Shin discloses that, when the distance between the user who provides the voice input and electronic device 101 is less than 1 meter, “the output amount of information may correspond to the most detailed original content corresponding to the voice input,” and when the distance between the user who provides the voice input and the electronic device 101 increases by 1 meter, “the output amount of information” decreases “by making the abbreviation level of the original content higher.” *Id.* ¶ 87.

Based on these cited disclosures, we find that Shin's processor 120 analyzes a user's speech (i.e., volume and speech rate) received as voice input from audio input module 151 to classify the user's voice based on distance thresholds between the user and electronic device 101. Amazon argues—and we agree—that Shin's Table 3 sets forth several levels of classification based on distance thresholds, one of which classifies the user's voice as a first voice when the distance between the user and electronic device 101 is less than 1 meter, and another that classifies the user's voice as a second voice when the distance between the user and electronic device 101 is more than 4 meters. *See* Pet. 14 (citing Ex. 1003, Fig. 3 (with Amazon's annotations of Figure 3 including the claimed “first voice” and “second voice” (in red))); Pet. Reply 11. As a result, we find that Shin discloses classifying a user's voice as either a first or second voice based on distance, as required by limitation 1[b].

We do not agree with SoundClear's argument that Shin only mentions voice or classification in the context of outputting content and, therefore, it does not disclose classifying voice input. *See* PO Resp. 24 (citing Ex. 1003 ¶ 51). It is well-settled that the test for anticipation “is not an ‘ipsissimis verbis’ test.” *In re Bond*, 910 F.2d 831, 832–33 (Fed. Cir. 1990). As we explain above, Shin's processor 120 effectively performs the voice classification required by limitation 1[b] because it analyzes a user's speech (i.e., volume and speech rate) received as voice input from audio input module 151 to classify the user's voice as either a first or second voice based on distance thresholds between the user and electronic device 101. This classification of the user's voice based on distance thresholds dictates the output scheme provided by electronic device 101 (e.g., starting with the

most detailed original content and ending with the most abbreviated or simplified version of the original content).

We also do not agree with SoundClear's characterization of Shin's Table 3 is being limited to a distance calculation and nothing else, nor do we agree that the deposition testimony of Amazon's declarant, Dr. Stern, supports SoundClear's characterization of Shin's Table 3. *See* PO Resp. 24–28; PO Sur-reply 6–8. Instead, we agree with Amazon's characterization of Shin's Table 3 as explicitly disclosing a classification scheme based on distance thresholds that is commensurate in scope with limitation 1[b]. *See* Pet. Reply 11. That is, the left-hand column of Shin's Figure 3 discloses levels of classification based on the distance between the user providing the voice input and electronic device 101 (i.e., distance is less than 1 meter, distance is greater than or equal to 1 meter but less than 2 meters, distance is greater than or equal to 2 meters but less than 3 meters, distance is greater than or equal to 3 meters but less than 4 meters, distance is greater than or equal to 4 meters), whereas the right-hand column of Shin's Figure 3 discloses the amount of output content corresponding to each classification level (i.e., most detailed original content, first abbreviation level, second abbreviation level, third abbreviation level, fourth abbreviation level/abstract content). *See* Ex. 1003 ¶¶ 86, 87.

ii. SoundClear's Second Argument

Second, SoundClear contends that limitation 1[b] requires the following: “(1) a classification that must be based on (2) an analysis of the acquired voice and (3) a distance, which is calculated (*e.g.*, by a proximity sensor).” PO Resp. 29 (citing Ex. 2013 ¶ 76). SoundClear argues that Shin does not disclose the claimed “[analyze/analyzing] the [acquired] voice to

classify the [acquired] voice” because Shin’s disclosure of processor 120 analyzing the voice input received through audio input module 151 “relates to *content* of speech”—not distance detection disclosure, let alone voice classification. *Id.* at 29–30 (citing Ex. 1003 ¶ 48; Ex. 2013 ¶ 77).

According to SoundClear, Shin’s paragraph 48 omits any discussion as to what is being analyzed and, instead, Amazon focuses on Shin’s output scheme, which SoundClear asserts is “an unrelated part of Shin’s disclosure.” *Id.* at 30. SoundClear then argues that Amazon entirely relies on the user-to-device distances set forth in Table 3 to disclose the claimed “classification,” but those distance detections are not based on any voice analysis. *Id.* at 30–31 (citing Ex. 1003, Table 3; Ex. 2013 ¶¶ 78, 39).

SoundClear, therefore, asserts that “there is no express disclosure or any other reason to believe Shin discloses ‘analyze [the] voice . . . to *classify the voice*.’” *Id.* at 31 (footnote omitted) (citing Ex. 2013 ¶ 79).

In its Reply, Amazon contends that “Shin discloses analyzing the acquired voice to classify it based on distance” because Shin explicitly discloses analyzing voice input by comparing it to user-to-distance threshold values to classify the voice into several different classifications so that the electronic device can generate the appropriate output (e.g., content, volume). Pet. Reply 14 (citing Ex. 1003 ¶¶ 37, 48, 76–80; Ex. 1202 ¶ 21). Amazon argues that Shin’s paragraph 48 is not just limited to processor 120 analyzing content of speech. *Id.* According to Amazon, SoundClear ignores Shin’s additional disclosures relating to analyzing a user’s voice. *Id.* (citing Pet. 12–13 (citing Ex. 1003, Abstract, ¶¶ 3–11, 37–39, 41, 47, 51, 60–62, 77–81, 89, 87, claims 1, 6, 15, 17, 20)). Amazon further argues that SoundClear’s arguments relating to determining user-to-device distance

based on voice analysis misses the mark because, even if limitation 1[b] were to require determining distance based on voice analysis, Shin discloses using a proximity sensor to calculate user-to-device distance, analyzing the voice input and determining the distance based on volume, or using both of these methods to more accurately determine distance, as evidenced by Shin's Figure 5A. *Id.* at 14–15 (citing Ex. 1003 ¶¶ 75–77, Fig. 5A).

In its Sur-reply, SoundClear contends that “Shin discloses only a limited form of acoustic analysis at the input stage,” which Shin calls a “voice analysis.” PO Sur-reply 11. SoundClear argues that, although Shin discloses receiving the speech of the user as a voice input to determine an output scheme, this is not sufficient to disclose “analyzing a voice input to *classify the voice*.” *Id.* at 12 (citing Pet. Reply 9). According to SoundClear, the claimed voice analysis is only accomplished “by using combinations of ‘feature values’ indicative of the characteristic of the voice.” *Id.* at 13. SoundClear admits that the specification of the '337 patent “exemplifies ‘distance’ as one such ‘feature values[]’ of many,” but SoundClear argues that “the plain language of the claims requires distance to be used in conjunction with other ‘feature values’ to classify voices.” *Id.* SoundClear asserts that the “foundational problem” with the Petition is that it “simply lacks any analysis as to how Shin classifies voices based on voice analysis.” *Id.* (citing Pet. 13–14).

Based on the complete record, we agree with Amazon that Shin discloses analyzing a user's voice to classify the user's voice, as required by limitation 1[b]. Pet. 11–14; Pet. Reply 9–14. As we explain above, we find that Shin's processor 120 analyzes a user's speech (i.e., volume and speech rate) received as voice input from audio input module 151 to classify the

user's voice based on distance thresholds between the user and electronic device 101. *See supra* Section II.B.4.a.1. We further find that Shin's Table 3 sets forth several levels of classification based on distance thresholds, one of which classifies the user's voice as a first voice when the distance between the user and electronic device 101 is less than 1 meter, and another that classifies the user's voice as a second voice when the distance between the user and electronic device 101 is more than 4 meters. *See supra id.*

We do not agree with SoundClear's argument that Shin's disclosure of processor 120 analyzing the voice input received through audio input module 151 only relates to analyzing speech content. *See* PO Resp. 29–30. Although SoundClear correctly notes that Shin's paragraph 48 discloses “processor 120 may analyze a voice input received through . . . audio input module 151,” we agree with Amazon that this paragraph in Shin should not be treated as limited to analyzing speech content. *See* Pet. Reply 14. Shin is replete with other disclosures indicating that its voice analysis includes more than just analyzing speech content. As one example, Shin discloses that “analysis of the voice input” includes obtaining “speech rate of a user's speech, a volume of a user's speech, and a keyword included in the user's speech.” Ex. 1003 ¶49. As another example, Shin discloses that “user 10 may speak ‘Let me know what time it is now.’ with a moderate volume and at a moderate speech rate, and the speech may be provided . . . as voice input.” *Id.* ¶37. As yet another example, Shin discloses that “user 10 may make a speech, such as ‘What time is it now?’, with a louder volume than usual and a faster speech rate than usual.” *Id.* ¶38. In each of these examples, Shin clearly demonstrates that processor 120 analyzes the speech of user 10 as voice input by analyzing speech content (e.g., “Let me know

what time it is now” or “What time is it now?”), together with the volume and speech rate (e.g., moderate or louder/faster).

We also do not agree with SoundClear’s argument that the plain language of the challenged claims requires distance to be used in conjunction with other feature values to classify a voice. *See* PO Sur-reply 13. As we explain in our claim construction analysis above, the plain language of claim phrase “analyze a voice . . . to classify the voice . . . based on distance between the user and the . . . device” controls because it merely requires analyzing a voice spoken by a user that is acquired by a device to classify the voice as either a first or second voice based on the distance between the user and the device. *See supra* Section II.A. Although the specification of the ’337 patent contemplates using feature values other than distance to perform the claimed analysis (e.g., intensity, peak frequency, slope of a peak in the spectrum), the plain language of the challenged claims only requires performing the claimed analysis using one feature value—specifically, distance. *See* Ex. 1001, 7:51–8:36.

In any event, even if we were to presume that the challenged claims require distance to be used in conjunction with other feature values to classify a voice, which, as we explain above, contravenes the plain language of the challenged claims, we agree with Amazon that Shin discloses this unduly narrow reading of limitation 1[b]. *See* Pet. Reply 14–15. That is, Shin’s processor 120 would still effectively perform the claimed analysis because it discloses analyzing a user’s voice to classify the user’s voice by analyzing a user’s speech received as voice input from audio input module 151 to classify the user’s voice based on both distance and volume. *See* Ex. 1003 ¶¶ 49, 53, 75–77, Fig. 5A, steps 501, 503.

b. Limitations 1[f] and 1[g]—“the voice-content generating unit is further configured to generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice,” and “generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice”

In its Patent Owner Response, SoundClear contends that, for the same reasons it identifies above with respect to limitation 1[b], Shin does not disclose classifying the acquired voice as a first and second voice, nor does Shin disclose generating a first and second output sentence when the acquired voice has been classified as a first and second voice. PO Resp. 32 (citing Ex. 2013 ¶ 81). In its Reply, Amazon contends that SoundClear’s argument that Shin does not disclose generating two different output sentences as required by limitations 1[f] and 1[g] is based on its voice classification arguments presented for limitation 1[b] and, therefore, fails for the same reasons. Pet. Reply 15–16.

Based on the complete record, we agree with Amazon that Shin discloses generating a first and second output sentence when the acquired voice has been classified as a first and second voice, as required by limitations 1[f] and 1[g]. Pet. 18–22; Pet. Reply 15. As we explain above, one example of the output scheme provided by processor 120 is illustrated in a consolidated view of Shin’s Table 3. This output scheme includes the following: when the distance between the user providing the voice input and electronic device 101 is less than 1 meter, “the output amount of information may correspond to the most detailed original content corresponding to the voice input,” and when the distance between the user providing the voice input and the electronic device 101 increases by 1 meter, “the output amount

of information” decreases “by making the abbreviation level of the original content higher.” *Id.* ¶ 87. Amazon argues—and we agree—this results in the following two classification levels: (1) when classifying the user’s voice as a first voice because the user is less than 1 meter from electronic device 101, processor 120 outputs the original content without abbreviation; and (2) when classifying the user’s voice as a second voice because the user is more than 4 meters from electronic device 101, processor 120 outputs abbreviated content by omitting information from the original content. *See* Pet. 19 (providing a consolidated, annotated version of Shin’s Table 3 with “First Voice” outlined in red), 21 (providing a consolidated, annotated version of Shin’s Table 3 with “Second Voice” outlined in red).

Apart from the arguments SoundClear presents for limitation 1[b], SoundClear does not address separately Amazon’s arguments and evidence that explain how Shin discloses limitations 1[f] and 1[g]. For the same reasons we identify above with respect to limitation 1[b], we do not agree with SoundClear’s arguments. *See supra* Section II.B.4.a.i and II.B.4.a.ii.

c. Limitation 1[h]—“the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto”

In its Patent Owner Response, SoundClear contends that Shin does not disclose limitation 1[h] because “[Amazon] relies on two separate tables disclosed in Shin: Table 1 (volume) and Table 3 (information).” PO Resp. 33 (citing Pet. 21–23). SoundClear argues that Amazon does not explain how the processes disclosed in connection with Shin’s “Tables 1 and 3 are compatible with one another,” nor does Amazon explain how a person of

ordinary skill in the art would have understood Shin's Tables 1 and 3 "to be used in conjunction with one another." *Id.*

In its Reply, Amazon contends that Shin's Tables 1 and 3 are compatible with one another because Shin explicitly discloses that the output volume and output information set forth in these two tables may be used together. Pet Reply 16 (citing Ex. 1003 ¶¶ 49, 53, 140, 183, 184, claim 17; Pet. 23; Ex. 1202 ¶ 23). In addition, relying on its declarant, Dr. Stern, Amazon asserts that a person of ordinary skill in the art would have considered "Shin's disclosure regarding output information and output volume together." *Id.* (quoting Ex. 1002 ¶ 88).

Based on the complete record, we agree with Amazon that Shin discloses an output controller that adjusts the volume of voice data, as required by limitation 1[h]. Pet. 22–24; Pet. Reply 15. In the context of describing Table 1, Shin discloses that "electronic device 101 may set an output volume level of the content to 40 dB when the distance between the user and . . . electronic device 101 is less than 1 m," and electronic device 101 "may increase the output volume level by 5 dB whenever the distance increases by 1 m." Ex. 1003 ¶ 80. According to Shin, this provides "a user who is far away from . . . electronic device 101 . . . with corresponding content [at] sound of a suitable output volume level." *Id.* Based on this cited disclosure, we find that, depending on the distance between a user providing voice input and electronic device 101, Shin discloses increasing the output volume level by 5 dB for every 1 meter the user moves away from electronic device 101.

When taking the aforementioned finding and applying it to the output scheme that includes the distance threshold classifications set forth in Shin's Table 3, Dr. Stern testifies that, when classifying a voice "as a first voice (e.g., closer than 1 meter), a first output sentence containing the 'original content' is provided (Table 3) at a volume level of 40 dB (Table 1)."

Ex. 1002 ¶ 88. Dr. Stern further testifies that, when classifying a voice "as a second voice (e.g., greater than 4 meters), a second output sentence containing '[c]ontent of fourth abbreviation level' is provided (Table 3) at a volume level of 60 or 65 dB (Table 1 and [paragraph 80])." *Id.* We credit Dr. Stern's testimony that Shin's Table 1 is intended to be used in conjunction with Shin's Table 3 for the following three reasons: (1) both Shin's Table 1 and 3 rely on the distance between the user and electronic device 101 to determine output volume level and output information, respectively; (2) Dr. Stern's testimony is consistent with other disclosures in Shin, such as while user 10 moves away from the electronic device, the electronic device "may provide the content as sound with a relatively louder output volume level" (Ex. 1003 ¶ 39); and (3) it is well-settled that "[w]hat a prior art reference discloses or teaches is determined from the perspective of one of ordinary skill in the art" (*Sundance, Inc. v. DeMonte Fabricating Ltd.*, 550 F.3d 1356, 1361 n.3 (Fed. Cir. 2008)).

We do not agree with SoundClear's argument that Shin's Tables 1 and 3 are not intended to be used in conjunction with one another. *See* PO Resp. 33. SoundClear's arguments in this regard fail to appreciate other disclosures in Shin that clearly contemplate that output volume and output content vary based on distance. *See, e.g.*, Ex. 1003 ¶¶ 39 (disclosing that, while user 10 moves away from electronic device 101, the electronic device

“may provide the content as sound with a relatively louder output volume level”), 49 (disclosing, “[f]or example,” that “the output scheme may include an output volume level . . . and an output amount of information of the content to be provided as sound”), 53 (disclosing that “processor 120 may determine a distance between the user and . . . electronic device 101 based on at least one of the volume of the user’s speech obtained through analysis of a voice input or distance computed, calculated, or measured by . . . distance detection module 180”). Moreover, SoundClear’s arguments do not appear to appreciate that Shin must be read through the perspective of a person of ordinary skill in the art, as evidenced by Dr. Stern’s testimony accompanying the Petition, which we credit above. *See* Ex. 1002 ¶ 88.

d. Shin discloses all the components as arranged in independent claim 1 for purposes of anticipation

In its Patent Owner Response, SoundClear contends that Amazon does not identify “where the components of [independent] claim 1 are actually disclosed in Shin,” but rather Amazon “focuses on functions without identifying what component allegedly accomplishes those functions.” PO Resp. 33 (citing Ex. 2013 ¶ 82). SoundClear asserts that Amazon “does not address the details or arrangement of the components/steps of any claim of the ’337 patent as required for anticipation.” *Id.* at 34.

In its Reply, Amazon contends that arguments and evidence with respect to Shin that are set forth in the Petition “consistently identif[y] the components accomplishing the claimed functions, which are the same components disclosed in the ’337 patent.” Pet. Reply 16 (citing Pet. 9 (proximity sensor), 11–18 (processor/CPU)).

Based on the complete record, we do not agree with SoundClear's argument that Amazon fails to explain how Shin accounts for the arrangement of components required by independent claim 1. *See* PO Resp. 33. In our view, Shin clearly and unequivocally discloses the claimed invention of independent claim 1, without any need for picking, choosing, and combining various disclosures in Shin not related directly to each other. *See In re Arkley*, 455 F.2d 586, 587 (CCPA 1972) (“[T]he [prior art] reference must clearly and unequivocally disclose the claimed [invention] or direct those skilled in the art to the [invention] without any need for picking, choosing, and combining various disclosures not directly related to each other by the teachings of the cited reference.”). For example, consistent with Amazon's claim mapping of independent claim 1, Shin's Figure 3 illustrates electronic device 101 that includes, among other things, processor 120, audio input module 151, audio output module 150, and distance detection module 180 (e.g., sensor) arranged in the same manner as the “proximity sensor,” “voice classifying unit,” “voice inquiring unit,” “process executing unit,” “voice-content generating unit,” and “output controller” required by independent claim 1. *See* Pet. 8–24. Notably, when arguing that Shin does not disclose all the components as arranged in independent claim 1 for purposes of anticipation, SoundClear does not identify, with particularity, a claimed component that is actually missing from Shin.

e. Remaining Arguments

SoundClear contends that Amazon's obviousness ground based on Shin is “cursory,” a “generalized attempt at a catchall,” “improperly conflates anticipation and obviousness” and, as a result, this obviousness ground “should be disregarded.” PO Resp. 35.

We do not agree with SoundClear’s argument that Amazon’s obviousness ground based on Shin suffers from the aforementioned deficiencies. As we explain above, we find Amazon presents sufficient evidence to support a finding that Shin discloses all the limitations of independent claims 1, 4, and 5, as arranged in these claims, and, therefore, Shin is an anticipatory reference. Nevertheless, we see no issue and, indeed, it is common practice, for a petitioner to offer alternative theories of unpatentability in a petition. The fact that Amazon makes alternative arguments in the Petition as to how Shin anticipates or renders obvious the challenged claims is not an improper assertion of unspecified grounds of unpatentability. *See* 35 U.S.C. § 312(a)(3) (requiring that a petition “identif[y], in writing and with particularity, each claim challenged[and] the grounds on which the challenge to each claim is based”).

SoundClear further argues that Amazon’s obviousness ground based on Shin “fails substantively” because Amazon’s explanation for modifying Shin’s Table 3 “would limit the [number] of distance-based *outputs*,” which, according to SoundClear, is contrary to the “classification *input* voices” required by the challenged claims. PO Resp. 35–36 (citing Ex. 2013 ¶ 84). SoundClear also argues that Amazon provides boilerplate language that amounts to bare and conclusory assertions to justify why a person of ordinary skill in the art would have modified Shin to arrive at the claimed inventions. *Id.* at 36.

As an initial matter, we do not understand Amazon to take the position that Shin’s Table 3 must be modified to account for certain limitations of independent claims 1, 4, and 5. Instead, as we explain above in our analysis of these independent claims, Amazon merely annotates the

consolidated view of Shin’s Table 3 to demonstrate that it discloses several levels of classification based on distance thresholds, one of which classifies the user’s voice as a first voice when the distance between the user and electronic device 101 is less than 1 meter, and another that classifies the user’s voice as a second voice when the distance between the user and electronic device 101 is more than 4 meters. *See* Pet. 14 (citing Ex. 1003, Fig. 3 (with Amazon’s annotations of the claimed “first voice” and “second voice” in red)); Pet. Reply 11.

Turning to SoundClear’s argument that Amazon provides boilerplate language that amounts to bare and conclusory assertions to justify why a person of ordinary skill in the art would have modified Shin to arrive at the claim inventions, there is no need to modify Shin in the manner proposed by Amazon. As we explain above, we find that Amazon presents sufficient evidence to support a finding that Shin discloses all the limitations of independent claims 1, 4, and 5, as arranged in these claims, and, therefore, Shin is an anticipatory reference.

f. Summary

In summary, Amazon has demonstrated by a preponderance of the evidence that the subject matter of independent claims 1, 4, and 5 is anticipated by Shin.

5. Claims 2 and 3

Claim 2 depends from independent claim 1, and further recites the following:

2[a] the process executing unit comprises:
an intention analyzing unit configured to extract intention information indicating an intention of the user based on the voice acquired by the voice acquiring unit; and

an acquisition content information acquiring unit configured to acquire acquisition content information which is notified to the user based on the extracted intention information, and

2[b] the voice-content generating unit is further configured to generate the text data including the acquisition content information as the output sentence.

Ex. 1001, 19:4–17 (with bracket annotations provided by Amazon in the Petition). Relying on the same arguments and evidence it presents for limitation 1[e], Amazon contends that Shin discloses limitation 2[a] because Shin’s processor 120 “analyzes the voice to identify the content the user is interested in, and then generates the requested content.” Pet. 25. Amazon argues that Shin’s electronic device may “receive the voice input . . . and may generate content corresponding to a result of analyzing the voice input.” *Id.* (quoting Ex. 1003 ¶ 37). As two examples, Amazon relies on Shin’s time example and weather example to demonstrate how the electronic device generates content in response to voice input received from a user. *Id.* (citing Ex. 1003 ¶¶ 39, 88). According to Amazon, Shin’s electronic device (e.g., electronic device 101) obtains this content because processor 120 includes an application that “analyz[es] a voice input[] and generat[es] and/or provid[es] the content corresponding to the voice input.” *Id.* (quoting Ex. 1003 ¶ 60). Amazon asserts that, because Shin’s electronic device is capable of delegating some processing to another device that delivers results, Shin effectively acquires information from memory or, alternatively, from another device. *Id.* (citing Ex. 1003 ¶¶ 64, 65, 67, 174–176; Ex. 1002 ¶ 93).

Amazon further contends that, even if Shin does not disclose explicitly all the functions of the claimed “process executing unit” required by limitation 2[a], Shin inherently discloses those functions because

processor 120 necessarily includes an “intention analyzing unit” and an “acquisition content information acquiring unit,” both of which would be required for Shin’s electronic device 101 to deliver the requested content. Pet. 25–26 (citing Ex. 1002 ¶ 94). For example, to provide the time and weather in Shin’s time and weather examples, processor 120 necessarily analyzes the voice input to determine the user’s intent (e.g., a request for time or weather) and necessarily requires corresponding information (e.g., “nine ten AM” or “cloudy”). *Id.* Lastly, Amazon contends that, even if Shin does not disclose explicitly or inherently all the functions of the claimed “process executing unit,” the functions of the claimed “process executing unit” required by limitation 2[a] still would have been obvious because the ’337 patent readily admits that those functions were known. *Id.* at 26 (citing Ex. 1001, 1:21–28; Ex. 1002 ¶ 96).

Amazon contends that Shin discloses limitation 2[b] because processor 120 acquires content (e.g., time or weather) when generating a response to the user. Pet. 27 (citing Ex. 1003 ¶¶ 37, 39, 88). According to Amazon, Shin provides this type of content to the user “as sound based on text to speech (TTS) technology.” *Id.* (quoting Ex. 1003 ¶ 48). Amazon, therefore, asserts that Shin’s processor 120 discloses the claimed “voice-content generating unit” because it generates the text data, including the acquisition content information (e.g., time or weather), as an output sentence. *Id.* (citing Ex. 1002 ¶ 98),

Claim 3 depends from independent claim 1, and further recites “wherein, on generating the second sentence, the voice-content generating unit is further configured to omit a part of information included in the voice spoken by the user.” Ex. 1001, 19:18–21. Relying on the same arguments

and evidence it presents for limitation 1[g], Amazon contends that Shin discloses this limitation because it classifies a voice received from a distance of more than 4 meters as the claimed “second voice.” Pet. 28. In the weather example identified previously, Amazon argues that Shin’s second output sentence is “rainy after cloudy” in response to the user’s request of “Let me know today’s weather.” *Id.* (citing Ex. 1003 ¶ 88). According to Amazon, Shin’s second output sentence in the weather example omits all the words spoken by the user, as well as information such as “Jul. 1, 2015.” *Id.* (citing Ex. 1003 ¶¶ 39, 88; Ex. 1002 ¶ 101). Amazon, therefore, asserts that Shin’s processor 120 generates the second output sentence by omitting part of the information included in the voice spoken by the user, as required by dependent claim 3. *Id.* (citing Ex. 1002 ¶ 101).

During trial, SoundClear asserts that Amazon “provides mere *ipse dixit*” with respect to dependent claim 2, and Amazon “provides curt, partial recitations of obviousness rationales” with respect to dependent claim 3 that are problematic, but even if complete, “they do not address the *claim* in full context.” PO Resp. 35–36. In our view, these bare assertions do not rise to the level of separately addressing Amazon’s arguments and evidence explaining how Shin discloses the additional limitation of dependents claims 2 and 3. We have reviewed Amazon’s arguments and supporting evidence explaining how Shin discloses the additional limitation of these dependent claims, and we agree with and adopt Amazon’s analysis. *See* Pet. 24–27. Accordingly, Amazon has demonstrated by a preponderance of the evidence that that the subject matter of dependent claims 2 and 3 is anticipated by Shin.

*C. Obviousness over Shimomura*⁴

Amazon contends that claims 1–5 of the ’337 patent are unpatentable under 35 U.S.C. § 103 as obvious over the teachings of Shimomura.

Pet. 35–59. Amazon contends that Shimomura teaches or suggests the subject matter of each challenged claim, and Amazon provides reasoning as to why a person of ordinary skill in the art would have been prompted to modify the teachings of this reference. *Id.* To support its positions, Amazon submits a Declaration of Dr. Stern with the Petition and a Reply Declaration of Dr. Stern with the Reply. *See* Exs. 1002, 1202.

SoundClear contends that the certified English translation of Shimomura is unreliable. PO Resp. 37–38. SoundClear further argues that Shimomura does not teach or suggest limitations 1[b], 1[f], 1[g], and 1[h]. PO Resp. 38–52; PO Sur-reply 5–11, 16–21. To support its positions, SoundClear submits the Declaration of Mr. McAlexander with the Patent Owner Response. *See* Ex. 2013.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1–5 of the ’337 patent are unpatentable as obvious over the teachings of Shimomura. We begin by

⁴ Current Office policy requires Board panels in an IPR to “address all grounds raised in the petition in [a] final written decision” because, among other reasons, “this approach will promote compact decision making and ensure that [Board] panels will not be required to address grounds for the first time in the event of a remand from a Director Review decision or Federal Circuit appeal.” *See* Memorandum with subject label “Final Written Decision Procedures for AIA Trial Proceedings” dated July 29, 2025, available at <https://www.uspto.gov/sites/default/files/documents/20250729095727582.pdf>.

providing an overview of Shimomura, and then we turn to the parties' contentions with respect to the challenged claims.

1. Overview of Shimomura

Shimomura is a Japanese Patent Application Publication. Ex. 1004, code (12). Shimomura generally relates “to a speech control device and method, and a robot device, which is suitable for application to, for example, an entertainment robot.” *Id.* ¶ 1. Shimomura discloses performance drawbacks associated with the equipment of conventional robots with speech recognition and interaction control functions, due to these conventional robots always conversing “with the user at a constant speech volume set in advance, regardless of the physical distance between the robot and the user.” *Id.* ¶ 4. Shimomura discloses that, depending on the speech volume setting of the robot, the volume may be too low for a user a little distance away from the robot to hear the speech content or, conversely, the volume may be too high for a user closer to the robot to hear the speech content. *Id.* ¶ 5. However, merely making it possible to freely change the speech volume of the robot impairs the naturalness of the interaction between the user and the robot, nor does it address when a room has a lot of reverberation. *Id.* ¶¶ 6–7.

To address these drawbacks, Shimomura discloses “improv[ing] the entertainment value of a robot” by providing “a dialogue control function.” Ex. 1004 ¶ 9. Shimomura discloses a “speech control device that controls the speech of the device during a dialogue between the device and a user” with “a speech form changing means . . . that changes the speech form of the device during a dialogue with the user as necessary according to the distance between the device and the user.” *Id.* ¶ 10. Figure 2 of Shimomura, reproduced below, illustrates “[a] block diagram showing the internal

[hardware] constitution of the robot” according to one embodiment. *Id.* ¶ 97.

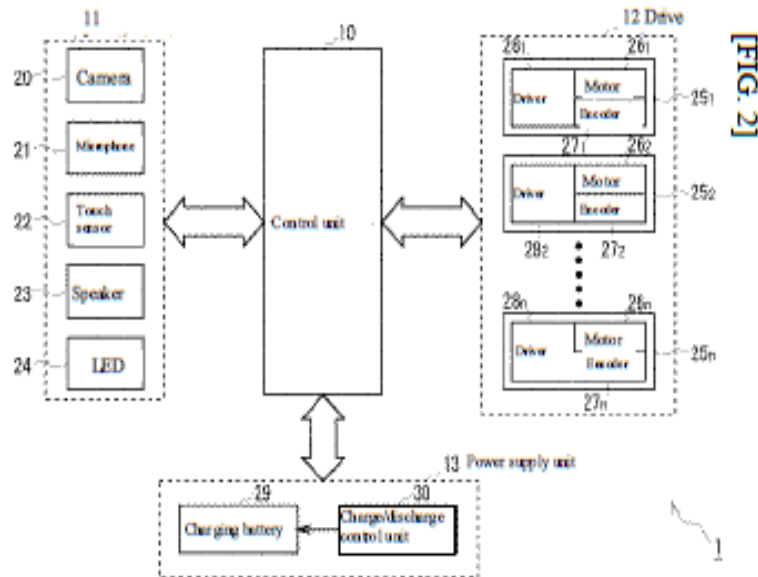


FIG. 2 Hardware constitution of the robot according to this embodiment (2)

Figure 2 illustrates control unit 10 for controlling the overall operation and data processing of the robot that is connected to input/output unit 11, drive unit 12, and power supply unit 13. *Id.* ¶ 20. Input/output unit 11 is further equipped with, among other things, cameras 20 corresponding to the robot’s eyes, microphones 21 corresponding to the robot’s ears, and speakers 23 that equates to the mouth of the robot. *Id.* ¶¶ 21, 22. The robot also includes software in the form of behavior control system 40 (not illustrated in Figure 2 above) that includes, among other things, image recognition unit 41 that “detects the distance to [an] imaging object” (*id.* ¶ 39), and speech recognition unit 42 that processes “speech recognition, speaker recognition, and sound source direction” by using speech signal S2 received from microphones 21 (*id.* ¶ 40).

2. English Translation of Shimomura

In the Petition, Amazon states that “Shimomura published as a Japanese Patent Application [Publication],” but for purposes of its obviousness grounds based, in whole or in part, on Shimomura, Amazon relies on the certified English translation of Shimomura. Pet. 6 (citing Ex. 1004). Amazon submits the Declaration of Gwen Snorteland with the Petition to support the accuracy of the certified English translation of Shimomura. Ex. 1005.

In its Patent Owner Response, SoundClear contends this IPR “turns on the precise disclosure of ‘voice,’ ‘analysis,’ and ‘classification,’” and “[t]here is no indication in the record that any particular consideration was given to the Japanese or English disclosures in Shimomura in that regard.” PO Resp. 27. SoundClear argues that “Ms. Snorteland has not laid the foundation for her understanding of the Japanese language (*e.g.*, fluency), let alone laid the foundation for her qualification as an expert on the distinction between ‘voice,’ and ‘distance’ or ‘volume’ in the context of the claimed ‘analy[sis] and classif[ication].” *Id.* at 38.

In its Reply, Amazon contends that the certified English translation of Shimomura accompanying the Petition is accurate and reliable. Pet. Reply 23. Amazon argues that, although “[SoundClear] questions the accuracy” of the certified English translation of Shimomura, “[SoundClear] did not submit its own translation; did not identify any inaccuracies in [Amazon’s] translation; did not object to its admissibility; did not depose [Amazon’s] translator; and did not dispute that the translator declaration ([Exhibit 1005]) attests to the translation’s accuracy.” *Id.* at 23. To avoid doubt, Amazon

submits an additional Declaration from Ms. Snorteland with the Reply “attesting to her fluency in Japanese.” *Id.* (citing Ex. 1206).

Based on the complete record, we determine that the certified English translation of Shimomura accompanying the Petition is both accurate and reliable. Our rule governing submission of evidence in an IPR states that, “[w]hen a party relies on a document . . . in a language other than English, a translation of the document into English and *an affidavit attesting to the accuracy of the translation must be filed with the document.*” 37 C.F.R. 42.63 (emphasis added). As an initial matter, we determine that the certifications set forth in the Declaration from Ms. Snorteland accompanying the Petition are compliant with § 42.63(b). Ex. 1005. We also see no reason to question Ms. Snorteland’s fluency in Japanese at the time she provided the certified English Translation of Shimomura because her testimony on this particular issue stands un rebutted. Ex. 1206 ¶ 2 (“As of September 5, 2024, and while preparing that declaration [accompanying the Petition], I was fluent in Japanese.”).

We do not agree with SoundClear’s argument that Ms. Snorteland must be an expert on the distinction between “voice,” “distance,” or “volume” as those terms are used in the challenged claims to provide an English translation of Shimomura. *See* PO Resp. 38. Amazon does not rely on Ms. Snorteland as an expert in the same field as the ’337 patent—Amazon relies on its declarant, Dr. Stern, for this purpose. *See* Exs. 1002, 1202. Amazon merely relies on Ms. Snorteland’s testimony to support the accuracy of the certified English translation of Shimomura that is Exhibit 1004. *See* Exs. 1005, 1206. Perhaps, more importantly, Amazon argues—and we agree—that there is no reason to question the accuracy of the

certified English translation of Shimomura provided by Ms. Snorteland because SoundClear did not submit its own English translation of Shimomura, SoundClear did not identify any inaccuracies in Ms. Snorteland's translation, and SoundClear did not take the opportunity to depose Ms. Snorteland. *See* Pet. Reply 23.

3. Claims 1, 4, and 5

Once again, our analysis below focuses on the limitations of independent claim 1 by relying on the bracketed annotations provided by Amazon in the Petition. *See* Pet. ii, 35–53. For the most part, the same arguments and evidence Amazon provides to explain how the teachings of Shimomura account for all the limitations of independent claim 1 equally apply to the corresponding limitations of independent claims 4 and 5. *See* Pet. 59.

Limitation 1[pre] recites “[a] voice-content control device comprising.” Ex. 1001, 18:39. To the extent the preamble is treated as limiting, Amazon contends that Shimomura teaches or suggests the features recited in the preamble because it discloses a “speech control device that controls the speech of a device” (e.g., a robot or other interactive device) by using input/output units, such as cameras 20, microphones 21, and speakers 23. Pet. 35–36 (quoting Ex. 1004, claim 1) (citing Ex. 1004, Title, ¶¶ 20–22, 96, Fig. 1; Ex. 1002 ¶ 130). Amazon argues that Shimomura's device “receives speech via microphones, performs speech recognition, and determines the ‘content to be spoken.’” *Id.* at 36 (quoting Ex. 1004 ¶ 46) (citing Ex. 1004, Abstract, Title, ¶¶ 1, 30, Fig. 4).

Limitation 1[a] recites “a proximity sensor configured to calculate a distance between a user and the voice-content control device.” Ex. 1001,

18:40–41. Amazon contends that Shimomura teaches or suggests this limitation because “[Shimomura’s] device includes a ‘distance detection means for detecting the distance to the conversation partner user.’” Pet. 37 (quoting Ex. 1004 ¶ 93). According to Amazon, this distance detection is achieved by using cameras 20 or a “distance sensor.” *Id.* (quoting Ex. 1004 ¶ 93) (citing Ex. 1004 ¶¶ 39, 59, 70, claims 1, 9, 17).

Limitation 1[b] recites “a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device.” Ex. 1001, 18:42–46. Amazon contends that Shimomura teaches or suggest the claimed “voice classifying unit” because its device (i.e., robot) includes control unit 10 and microphones 21. Pet. 37–38 (citing Ex. 1004, Fig. 2; Ex. 1002 ¶ 137). Amazon argues that Shimomura’s control unit 10, which includes a CPU that controls overall operation and data processing, implements a “behavior control system [40]” that further includes “speech recognition unit [42].” *Id.* (first quoting Ex. 1004 ¶ 36, then quoting Ex. 1004 ¶ 38) (citing Ex. 1004 ¶¶ 20, 27, 28, 32, 34, 37, Fig. 3). Amazon argues that Shimomura teaches or suggests the claimed “voice acquiring unit” because “speech recognition unit 42 performs various sound-related recognition processes such a speech recognition, speaker recognition, and sound source recognition based on the speech signal S2 provided by . . . microphone[s] 21.” *Id.* at 38–39 (quoting Ex. 1004 ¶ 40).

Amazon further contends that Shimomura’s control unit 10 also includes “speech synthesis unit [48]” and “speech character string transformation unit [51].” Pet. 39 (quoting Ex. 1004 ¶ 67) (citing Ex. 1004

¶¶ 36, 37, 46, 58, Fig. 4; Ex. 1002 ¶ 138). According to Amazon, these two units in Shimomura may “transform the character string” of the content spoken by the robot or change its volume “according to the distance to the conversation partner.” *Id.* (quoting Ex. 1004 ¶ 58) (citing Ex. 1004 ¶¶ 60–63, 67–70, 73–84, Abstract, claims 1–4). Amazon asserts that, prior to accomplishing this character string transformation and volume change, Shimomura’s control unit 10 first classifies the voice into the following two classifications based on the user-to-device distance: (1) a first voice at a distance more than 350 centimeter (“cm”); and (2) a second voice at a distance less than 350 cm. *Id.* (citing Ex. 1002 ¶¶ 138, 139). Amazon notes that the two distance classifications are represented in the annotated version of Shimomura’s Figure 7, reproduced below.

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

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

Id. at 39–40. The annotated version of Figure 7 illustrates a character string transformation table with the first voice at a distance more than 350 cm from the device (in red) and the second voice at a distance less than 350 cm from the device (also in red). *Id.* 40 (citing Ex. 1004 ¶¶ 60–64, Fig. 7; Ex. 1002 ¶ 139).

Limitation 1[c] recites “a process executing unit configured to analyze the voice acquired by the voice acquiring unit to execute processing required by the user.” Ex. 1001, 18:47–49. Relying on the same arguments and evidence it presents for limitation 1[b], Amazon contends that Shimomura’s control unit 10 teaches or suggests the claimed “process executing unit” because it analyzes the voice acquired by speech recognition unit 42 by recognizing the speech from the signal emanating from microphones 21 and, based on that speech, determines the content to be spoken to the user. Pet. 40 (citing Ex. 1004 ¶ 40; Ex. 1002 ¶ 142). Amazon argues that, because Shimomura describes the user as a conversation partner, this confirms that its device extracts the user intent during a conversation and responds by performing the required processing. *Id.* (citing Ex. 1004 ¶¶ 3, 4, 50, 54, 57–59, 67, 86, 87, 90, claims 1, 9, 17; Ex. 1002 ¶ 142).

Limitation 1[d] recites “a voice-content generating unit configured to generate, based on content of the processing executed by the processing executing unit, output sentence that is text data for a voice to be output to the user.” Ex. 1001, 18:50–53. Relying on the same arguments and evidence it presents for limitations 1[b] and 1[c], Amazon contends that Shimomura’s control unit 10 teaches or suggests the claimed “voice-generating unit” because it generates content for output (e.g., text data like “Today’s weather in Tokyo is sunny”). Pet. 41 (quoting Ex. 1004 ¶ 65) (citing Ex. 1004 ¶¶ 76, 78, 79, 92). Amazon argues that Shimomura’s control unit 10 also includes both “output management unit [46] and “behavior selection control unit [45],” where output management unit 46 “sends the character string provided by the behavioral selection control unit 45 to the speech synthesis unit 48, while the speech synthesis unit 48 generates a speech signal S4 of

synthetic speech based on the supplied character string and sends it to the speaker[s] 23” for audible presentation. *Id.* at 41–42 (first quoting Ex. 1004 ¶ 46, then quoting Ex. 1004 ¶ 47) (citing Ex. 1004 ¶¶ 36, 37, 49, 50, 63, 83, claims 6, 4, Figs. 4, 13; Ex. 1002 ¶ 146). Amazon further argues that Shimomura’s control unit 10 generates an output sentence “based on content of the processing executed by the processing executing unit” because it determines the content spoken to the user based on recognized speech so that the device converses with the user. *Id.* at 42 (citing Ex. 1002 ¶ 147).

Limitation 1[e] recites “an output controller configured to adjust a sound volume of voice data obtained by converting the output sentence thereinto.” Ex. 1001, 18:54–56. Relying on the same arguments and evidence it presents for limitation 1[d], Amazon contends that Shimomura teaches or suggest the claimed “output controller” because control unit 10 generates the speech signal and sends it to speakers 23. Pet. 42. Amazon also argues that Shimomura’s control unit 10 implements speech synthesis unit 48, which stores in memory “a volume change amount specification table 52 that specifies the amount of change in the speech volume of the robot 1 with respect to distance to the conversation partner.” *Id.* at 42–43 (quoting Ex. 1004 ¶ 67) (citing Ex. 1004 ¶¶ 50, 68–73, 75, 83, 84, 90, 92, claims 2, 10, 18, Figs. 9, 13). Amazon then asserts that Shimomura’s speech synthesis unit 48 determines the volume output level based on distance thresholds between the robot and the conversation partner. *See id.* at 43 (citing Ex. 1004 ¶ 70).

Limitation 1[f] recites “the voice-content generating unit is further configured to generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice.” Ex. 1001, 18:57–

60. Amazon contends that Shimomura teaches or suggests this limitation because control unit 10, which implements speech character string transformation unit 49, inserts a word like “ne” after each phrase to “chang[e] the interval between phrases so that the division of meaning can be clearly understood according to the physical distance to the conversation partner.” Pet. 43 (quoting Ex. 1004 ¶ 57) (citing Ex. 1004 ¶¶ 58–66, 85, 92, Figs. 7, 8). According to Amazon, when Shimomura classifies the voice as a first voice at a distance more than 350 cm, the device generates a first output by transforming the original content to appropriately insert “ne” or another word after each phrase. *Id.* at 43–44 (citing Ex. 1004 ¶¶ 62, 65, 92, Fig. 7; Ex. 1002 ¶ 152). For example, Amazon argues that Shimomura disclose “generat[ing] a character string such a ‘Today’s, Tokyo’s, the weather is, sunny’ by adding the character ‘ne’ to the end of each clause-final auxiliary verb, ‘Today’s,’ ‘Tokyo’s,’ and ‘The weather is.’” *Id.* at 44 (quoting Ex. 1004 ¶ 65) (citing Ex. 1002 ¶ 153).

Limitation 1[g] recites the voice-content generating unit is further configured to “generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as a second voice.” Ex. 1001, 18:61–64. Amazon contends that Shimomura teaches or suggest this limitation because speech character string transformation unit 51 “judges that the character string should not be transformed if the distance to the conversation partner . . . is less than 350 [cm].” Pet. 45 (quoting Ex. 1004 ¶ 62) (citing Ex. 1004 ¶¶ 57, 58, 65, 85, 92). According to Amazon, when Shimomura classifies the voice as a second voice at a distance less than 350 cm, the device generates a second output comprising the original content without

transformation. *Id.* (citing Ex. 1004 ¶¶ 62, 92, Fig. 7; Ex. 1002 ¶ 157). For example, Amazon argues that Shimomura discloses that, when classifying the voice as a second voice (e.g., less than 350 cm), the output sentence may be: “Today’s weather in Tokyo is sunny.” *Id.* at 46 (quoting Ex. 1004 ¶ 65). In this weather example, Amazon asserts that Shimomura’s output sentence omits “ne” or another word after each phrase. *Id.* (citing Ex. 1002 ¶ 158).

Limitation 1[h] recites “the output controller is further configured to adjust the sound volume of the voice data such that the sound volume of the voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto.” Ex. 1001, 18:65–19:3. Relying on the same arguments and evidence it presents for limitation 1[e], Amazon contends that Shimomura’s control unit 10 adjusts the output volume level of the content based on the user-to-device distance. Pet. 46. For example, Amazon argues that control unit 10 implements speech synthesis unit 48, which stores in memory the volume change table illustrated in Figure 9. *Id.* (citing Ex. 1004 ¶ 70, Fig. 9; Ex. 1002 ¶ 161). Amazon further contends that, although Shimomura generally discloses adjusting the volume level based on distance, it does not specify that different output sentences have different volumes. *Id.* at 47. Amazon asserts that modifying Shimomura in this way would have been obvious to a person of ordinary skill in the art. *Id.* (citing Ex. 1002 ¶¶ 162–172).

More specifically, Amazon modifies Shimomura’s volume change table illustrated in Figure 9 as follows: (1) by extending the 150–250 cm range to end at 350 cm, thereby increasing the last range to more than 350

cm or, alternatively, (2) by adding a new range from 250–350 cm, thereby increasing the last range to more than 350 cm. Pet. 47 (citing Ex. 1002 ¶ 163). According to Amazon, these two modifications are illustrated in the first and second modified volume change tables, reproduced below.

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

First Modified Volume Change Table

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

Second Modified Volume Change Table

Id. Amazon’s first modified volume change table illustrates the extension of the 150–250 cm range to end at 350 cm (in yellow), and Amazon’s second modified volume change table illustrates adding a new range from 250–350 cm and increasing the last range to more than 350 cm (both in yellow). *Id.*

Amazon contends that, when considering the first modified volume change table above together with the character string transformation table illustrated in Figure 7, it would have been obvious to a person of ordinary skill in the art to provide two classification levels for a user’s voice. Pet. 50 (citing Ex. 1002 ¶ 171). Amazon argues that, when classifying the voice as a first voice at a distance more than 350 cm, Shimomura would provide a first output sentence containing the original content after transformation (i.e., by including the word “ne” or another word after each phrase) with a volume level +5 higher than the reference volume level. *Id.* Amazon further argues that, when classifying the voice as a second voice at a

distance less than 350 cm, Shimomura would provide a second output sentence containing the original content without transformation (i.e., the word “ne” or another word is not included after each phrase) with a volume level of +3 higher than the reference volume level. *Id.* According to Amazon, its first modified volume change table, reproduced below, represents these two classification levels for the user’s voice.

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

First Modified Volume Change Table

Id. at 51. Amazon’s first modified volume change table illustrates the first voice at a distance of more than 350 cm with a volume level of +5 higher than the reference volume level (in red), and the second voice at a distance between 150–350 cm with a volume level of +3 higher than the reference volume level (also in red). *Id.*

Amazon provides six reasons as to why a person of ordinary skill in the art would have been motivated to modify Shimomura in this way. Pet. 48–50. As one example, Amazon argues that such modifications would have been a simple substitution of a known element (i.e., a 350 cm distance threshold as illustrated in Shimomura’s Figures 7, 10, and 11) for another known element (i.e., a 250 cm distance threshold as illustrated in Shimomura’s Figure 9) to obtain predictable results (i.e., a device that adjusts output volume based on whether a user is less than or more than 350 cm from the device). *Id.* at 48 (citing Ex. 1002 ¶ 164). Amazon further contends that a person of ordinary skill in the art would have had a

reasonable expectation of success in modifying Shimomura's volume change table illustrated in Figure 9 because doing so would have been a "trivial modification" involving "configuring Shimomura's control unit 10 with the revised table[s] to provide volume levels based on whether the user is [less than] or [more than] 350 cm" from the device. *Id.* at 50 (citing Ex. 1002 ¶ 170).

Notwithstanding SoundClear's arguments to the contrary, Amazon has shown by a preponderance of the evidence that Shimomura teaches or suggests all the limitations of independent claims 1, 4, and 5. During trial, SoundClear contends that Shimomura does not teach or suggest limitations 1[b], 1[f], 1[g], and 1[h]. PO Resp. 38–52; PO Sur-reply 5–11, 16–21. We address SoundClear's arguments in turn.

a. Limitation 1[b]—"a voice classifying unit configured to analyze a voice spoken by a user and acquired by a voice acquiring unit to classify the voice as either one of a first voice or a second voice based on the distance between the user and the voice-content control device"

In its Patent Owner Response, SoundClear presents the following two arguments as to why Shimomura does not teach or suggest limitation 1[b]: (1) Shimomura discloses measurement of distance—not classification of voices (PO Resp. 38–42; PO Sur-reply 5–11); and (2) Shimomura does not disclose analysis of a voice for classification (PO Resp. 42–44; PO Sur-reply 16–18). Although these two arguments overlap in certain respects, we address them in turn.

i. SoundClear's First Argument

First, SoundClear contends that Shimomura "does not disclose any manner of classifying voices." PO Resp. 38 (citing Ex. 2013 ¶ 86). According to SoundClear, the only basis Amazon relies on to teach or

suggest classification of a voice is Shimomura's Figure 7, but Amazon does not identify an explicit disclosure of "classification" or "voice" in the corresponding description of this figure. *Id.* at 38–39. SoundClear argues that Amazon also does not provide a reasoned analysis as to why a person of ordinary skill would equate measurement of distance with either "classification" or "voice." *Id.* at 39. SoundClear argues that, to the extent Shimomura teaches the classification required by limitation 1[b], it only teaches classes of distances because Shimomura's Figure 7 merely illustrates "a calculation of distance and nothing else." *Id.* at 39–40 (citing Ex. 2013 ¶¶ 87, 88); *see also id.* at 41 (arguing the same). To support its argument that Shimomura draws a distinction between "voice," "distance," and "volume," SoundClear directs us to the background disclosure in Shimomura regarding conventional robots conversing at a constant speech volume, regardless of distance. *Id.* at 40 (citing Ex. 1004 ¶ 4; Ex. 2013 ¶ 89). In short, SoundClear asserts that, "like Shin, Shimomura directly determines the output content and volume based on distance without performing the step of classifying the voice." *Id.* at 41 (citing Ex. 2013 ¶ 91).

In its Reply, Amazon contends that as a matter of law Shimomura is not required to use the exact words of the challenged claims, but there can be no dispute Shimomura analyzes a voice by receiving speech input and then classifies it based on distance, which is all that limitation 1[b] requires. Pet. Reply 18 (citing Pet. 38–40). Amazon argues that, contrary to SoundClear's assertions, Shimomura's Figure 7 teaches more than calculation of distance—this figure teaches classification of voice input based on whether a user's distance exceeds a threshold of 350 cm. *Id.* Amazon argues that

SoundClear ignores that Shimomura's Figure 7 teaches "classes or categories of distances *from which the voice is received*" because "[t]he top row in Figure 7 includes two discrete classifications into which the voice may fall" and "the bottom row [in Figure 7] describes the output result of each classification." *Id.* at 18–19 (footnote omitted) (citing Ex. 1004, Fig. 7; Ex. 1201, 112:3–113:2; Ex. 1202 ¶ 24).

In its Sur-reply, SoundClear contends that Amazon does not identify how Shimomura teaches or suggests classifying voice input, but rather both Amazon and its declarant, Dr. Stern, improperly rely on unsupported assumptions based on Shimomura's inputs and outputs. PO Sur-reply 5–6; *see also id.* at 9 (arguing that Amazon "identif[ies] no express disclosure of any classification of inputs in . . . Shimomura"). According to SoundClear, Dr. Stern provides deposition testimony confirming that adjusting an output based on an input does not disclose classifying voice input. *Id.* at 6–8 (citing Ex. 2016, 21:8–22:16, 33:5–34:15, 62:22–63:8, 64:5–65:2, 65:7–12, 77:19–78:3). Based on the cited portions of Dr. Stern's deposition testimony, SoundClear argues that a person of ordinary skill in the art would not assume that classification necessarily occurs in Shimomura. *See id.* at 8.

Based on the complete record, we agree with Amazon that Shimomura teaches or suggests classifying a user's voice as either a first or second voice based on distance thresholds between the user and the device, as required by limitation 1[b]. Pet. 37–40; Pet. Reply 18–19. Shimomura's robot includes "control unit 10 for controlling overall operation and other data processing" (Ex. 1004 ¶ 20, Fig. 2) that is configured with "CPU . . . 31 as the main controller, which is connected to memory, other circuit components and peripheral devices via a bus" (*id.* ¶ 27, Fig. 3).

The memory of the robot stores software in the form of behavior control system 40, which includes, among things, speech recognition unit 42, behavior control system 45, and speech synthesis unit 48. *Id.* ¶¶ 36–47, Fig. 4.

Shimomura discloses that “speech recognition unit 42 performs various sound-related recognition processes such as speech recognition, speaker recognition, and sound source direction recognition based on the speech signal S2 provided by . . . microphone[s] 21.” Ex. 1004 ¶ 40. Speech synthesis unit 48 works in conjunction with speech character string transformation unit 51 to transform the content spoken to the robot by “inserting ‘ne’ after each phrase according to the distance to the conversation partner.” *Id.* ¶¶ 58, 68, Fig. 7. One example of this transformation process is illustrated in Figure 7 of Shimomura, reproduced below.

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

Figure 7 of Shimomura illustrates that this transformation process includes two classifications based on the distance between the robot and the conversation partner user: (1) if the voice of the user is at a distance more than 350 cm from the robot, the voice is transformed by inserting “ne” after each phrase; and (2) if the voice of the user is at a distance of less than 350

cm from the robot, the voice is not transformed and the “ne” is omitted after each phrase. *Id.* ¶¶ 60–64, Fig. 7.

Based on these cited disclosures, we find that Shimomura’s control unit 10 employs software that analyzes the speech of a conversation partner user received as voice input from microphones 21 to classify the user’s voice based on distance thresholds between the user and the robot. Amazon argues—and we agree—that Shimomura’s Figure 7 sets forth two levels of classification based on distance thresholds, one of which classifies the user’s voice as a first voice when the distance between the user and the robot is more than 350 cm, and another that classifies the user’s voice as a second voice when the distance between the user and the robot is less than 350 cm. *See* Pet. 39–40 (citing Ex. 1004, Fig. 7 (with Amazon’s annotations of Figure 7 including the claimed “first voice” and “second voice” (in red))); Pet. Reply 18–19. As a result, we find that Shimomura teaches or suggests classifying a user’s voice as either a first or second voice based on distance thresholds between the user and the device, as required by limitation 1[b].

We do not agree with SoundClear’s argument that the corresponding description of Shimomura’s Figure 7 does not mention voice or classification in the context of outputting content and, therefore, it does not teach or suggest classifying a voice. *See* PO Resp. 38–39. When conducting an obviousness analysis, we must consider Shimomura “not only for what it expressly teaches, but also for what it fairly suggests.” *Bradium Techs. LLC v. Iancu*, 923 F.3d 1032, 1049 (Fed. Cir. 2019). As we explain above, Shimomura’s control unit 10 implements software that teaches or suggests performing the voice classification required by limitation 1[b] because it analyzes the speech of a conversation partner user received as

voice input from microphones 21 to classify the user's voice as either a first or second voice based on distance thresholds between the user and the robot. This classification of the user's voice based on distance thresholds dictates the output scheme provided by the robot (e.g., starting with the original content without character string transformation and ending with transforming the original content by adding character strings).

We also do not agree with SoundClear's characterization of Shimomura's Figure 7 as being limited to a distance calculation and nothing else, nor do we agree that the deposition testimony of Amazon's declarant, Dr. Stern, supports SoundClear's characterization of Shimomura's Figure 7. *See* PO Resp. 39–41; PO Sur-reply 6–8. Instead, we agree with Amazon's characterization of Shimomura's Figure 7 as explicitly disclosing a classification scheme based on distance thresholds that is commensurate in scope with limitation 1[b]. *See* Pet. 39–40; Pet. Reply 18–19. That is, the top row of Shimomura's Figure 7 discloses two levels of classification based on the distance between the conversation partner user and the robot (i.e., distance is less than 350 cm or distance of more than 350 cm), whereas the bottom row of Shimomura's Figure 7 discloses the output amount of content corresponding to each classification level (i.e., do not implement the character string transformation or implement character string transformation). *See* Ex. 1004 ¶¶ 60–64, Fig 7.

SoundClear's reliance on the background disclosure in Shimomura regarding conventional robots conversing at a constant speech volume, regardless of distance, is misplaced. PO Resp. 40 (citing Ex. 1004 ¶ 4). This background disclosure appears under the header "Problem to be Solved by the Invention." This background disclosure identifies drawbacks

associated with “conventional robots equipped with . . . voice recognition and dialogue control functions” (Ex. 1004 ¶ 4) that Shimomura seeks to address by “improv[ing] the entertainment value of a robot having a dialogue control function” (*id.* ¶ 9). Stated differently, SoundClear improperly focuses on the problem addressed by Shimomura rather than its proposed solution.

ii. SoundClear’s Second Argument

Second, SoundClear contends that limitation 1[b] requires the following: “(1) analysis of the acquired voice; and (2) distance separately calculated (*e.g.*, by a proximity sensor).” PO Resp. 42 (citing Ex. 2013 ¶ 94). SoundClear argues that Shimomura does not teach or suggest the claimed “[analyze/analyzing] the [acquired] voice to classify the [acquired] voice” because—like Shin—Amazon does not present sufficient evidence to support a finding that Shimomura’s control unit 10 implements software that analyzes the voice input received through microphones 21 that actually determines user-to-device distance based on “any analysis of the user’s voice.” *Id.* at 42–43 (citing Pet. 37–40; Ex. 2013 ¶ 95). That is, SoundClear asserts that Shimomura does not teach or suggest “determining user-to-device distance based on any voice analysis,” but rather “Shimomura relies exclusively on camera 20 in image recognition unit 41 to detect distance.” *Id.* at 43–44 (citing Ex. 2015, 45:3–17, 53:17–22, 54:1–6; Ex. 2013 ¶ 95).

In its Reply, Amazon contends SoundClear’s argument that Shimomura’s distance calculation does not “involve[] any analysis of the user’s voice” because Shimomura uses camera 20 to detect distance “is irrelevant.” Pet. Reply 19 (citing PO Resp. 43–44). Amazon argues that nothing in limitation 1[b] requires distance to be based on voice analysis, nor

does the '337 patent describe determining distance based on any analysis of the voice signal—the '337 patent only describes determining distance using a proximity sensor. *Id.* at 19–20 (citing Ex. 1001, 8:20–34). Amazon asserts that Shimomura teaches or suggests “calculating the user-to-device distance in the same way as the '337 patent” because it discloses classifying a voice based on whether the distance exceeds a threshold. *Id.* at 20 (citing Ex. 1001, 8:20–34, claim 1; Ex. 1004 ¶ 93; Pet. 37–40; Ex. 1202 ¶ 26).

In its Sur-reply, SoundClear contends that “[Amazon does] not and cannot dispute that Shimomura only discloses detecting distance using a camera.” PO Sur-reply 16 (citing Pet. Reply 19–20). SoundClear argues that limitation 1[b] requires “analyz[ing] a voice . . . to classify the voice . . . based on the distance,” but the distance detection in Shimomura is different because “Figure 7 is only used to change the speech output of Shimomura’s robot.” *Id.* at 16–17 (citing Ex. 1004 ¶¶ 57, 58). SoundClear, therefore, asserts that the obviousness ground based on Shimomura alone should fail because “a [person of ordinary skill in the art] would not have understood Shimomura’s classes of distance (described in connection with Figure 7) as disclosing the claimed voice classification.” *Id.* at 17 (citing PO Resp. 38–42).

Based on the complete record, we agree with Amazon that Shimomura teaches or suggests analyzing a user’s voice to classify the user’s voice, as required by limitation 1[b]. Pet. 37–40; Pet. Reply 18–19. As we explain above, we find that Shimomura’s control unit 10 employs software that analyzes the speech of a conversation partner user received as voice input from microphones 21 to classify the user’s voice based on distance thresholds between the user and the robot. *See supra* Section

II.C.3.a.i. We further find that Shimomura's Figure 7 sets forth several levels of classification based on distance thresholds, one of which classifies the user's voice as a first voice when the distance between the user and robot is more than 350 cm, and another that classifies the user's voice as a second voice when the distance between the user and the robot is less than 350 cm. *See supra id.*

We do not agree with SoundClear's argument that Shimomura does not determine user-to-device distance based any voice analysis because this argument is predicated on the notion that Shimomura relies exclusively on cameras 20 to detect distance. *See* PO Resp. 43–44; PO Sur-reply 16. This is incorrect. Although SoundClear correctly notes that Shimomura's image recognition unit 41 “detects the distance to the imaging object by means of a so-called stereo visual method based on the image signal S1 from [charged coupled device (“CCD”)] camera 20,” Shimomura makes clear that its distance detection is not so limiting. Ex. 1004 ¶ 39. Shimomura discloses that, “in the above embodiment, the distance detection means for detecting the distance to the conversation partner user is configured by . . . CCD cameras 20 . . . , but the present invention is not limited to this, and other means such as a distance sensor can be widely applied.” *Id.* ¶ 93. Based on these cited disclosures, we find that Shimomura is not limited to detecting distance using cameras 20, but clearly contemplates using other means for detecting distance, such as using a distance sensor.

In any event, how Shimomura detects distance between the conversation partner user and the robot is only one piece of the puzzle. As we explain above, Shimomura uses this distance detection between the user and the robot to determine a distance threshold, thereby enabling control unit

10 that employs software for analyzing the speech of the conversation partner user received as voice input from microphones 21 to determine the proper classification (i.e., a first voice when the distance between user and the robot is more than 350 cm, or a second voice when the distance between the user and the robot is less than 350 cm). Put simply, Shimomura's distance detection is tied inextricably to analyzing the speech of the conversation partner user received as voice input from microphones 21 for purposes of classifying it as a first or second voice.

We also do not agree with SoundClear's argument that Shimomura's Figure 7 is only used to change the speech output of the robot. *See* PO Sur-reply 16–17. As an initial matter, this argument is internally inconsistent with SoundClear's argument that Shimomura's Figure 7 should be limited to a distance calculation and nothing else. *Compare* PO Resp. 39–41, *with* PO Sur-reply 16–17. In any event, for the same reasons we already identify above, we do not agree with SoundClear's characterization of Shimomura's Figure 7. *See supra* Section II.C.3.a.i.

b. Limitations 1[f] and 1[g]—“the voice-content generating unit is further configured to generate a first output sentence as the output sentence when the acquired voice has been classified as the first voice,” and “generate a second output sentence in which information is omitted as compared to the first output sentence as the output sentence when the acquired voice has been classified as the second voice”

SoundClear contends that Shimomura generates a second output sentence by omitting information that is disclosed in the form of “ne” after each phrase. PO Resp. 51 (citing Ex. 2013 ¶ 107). SoundClear, however, asserts that the relevant disclosure in Shimomura indicates that the omission

of “ne” is a “change[] in intonation” and “not an omission of information.” *Id.* (quoting Ex. 1004 ¶ 92).

In its Reply, Amazon contends that “Shimomura does not specify that it is talking about ‘ne’ when it discusses change in intonation.” Pet. Reply 22–23 (citing Ex. 1004 ¶ 92). According to Amazon, “Shimomura’s disclosure is not limited to what [SoundClear] argues” because Shimomura discloses inserting words other than “ne” after each phrase. *Id.* at 23 (citing Ex. 1004 ¶ 92; Ex. 1202 ¶ 28).

Based on the complete record, we agree with Amazon that Shimomura teaches or suggests generating two different output sentences based on classifying a user’s voice as first or second voice, as required by limitations 1[f] and 1[g]. Pet. 43–46; Pet. Reply 23. The parties’ dispute regarding limitations 1[f] and 1[g] turns on whether the Japanese word “ne” is merely a change in intonation, such that its absence would not amount to an omission of information. As an initial matter, the parties do not direct us to evidence in the record as to the English translation of the Japanese word “ne.” Despite the lack of evidence to demonstrate the meaning of the Japanese word “ne” in the record before us, Shimomura discloses that “[i]t goes without saying that words other than ‘ne’ can be used as words to be inserted after each phrase end to create a gap between phrases.” Ex. 1004 ¶ 95.

Based on this cited disclosure, we find that the Shimomura’s character string transformation that occurs when the conversation partner user is more than 350 cm from the robot includes inserting words (whatever those words may be) after each phrase. Whether those words also change intonation is of no moment because Shimomura clearly refers to them as “words.” Patent

Owner also appears to acknowledge that “phrases” like “ne” are potentially “undesirable” or “superfluous” additional information that can be added or omitted. PO Resp. 53 (“In the context of Shimomura, however, a [person of ordinary skill in the art] would have recognized that doing so would have been undesirable as it would have burdened Shimomura’s robot and user with superfluous phrases (e.g., ‘ne’) when they are unnecessary for the robot to be understood.”). Based on the character string transformation determination illustrated in Shimomura’s Figure 7, those words (once again, whatever those words may be) would be omitted when the conversation partner user is less than 350 cm away from the robot, thereby resulting in a second output sentence in which information is omitted as compared to the first output sentence.

- c. *Limitation 1[h]— “the output controller is further configured to adjust the sound volume of voice data such that the sound volume of voice data obtained by converting the first output sentence thereinto differs from the sound volume of the voice data obtained by converting the second output sentence thereinto”*

In its Patent Owner Response, SoundClear contends that “[Amazon] does not provide any analysis supporting the assertion that the processes disclosed in connection with [Shimomura’s] Figures 7 and 9 are compatible with one another,” nor does Amazon explain “how a [person of ordinary skill in the art] would have understood them to be used in conjunction with one another.” PO Resp. 45. According to SoundClear, Shimomura’s Figures 7 and 9 are incompatible because the distances associated with these figures are different. *Id.* at 45–46 (citing Ex. 2013 ¶¶ 98–100). SoundClear further argues that a person of ordinary skill in the would have had no reason to modify the distances associated with the volume change table in

Shimomura's Figure 9 to align with the distances associated with character string transformation determination table in Shimomura's Figure 7 because each table is "tied to separate 'processing procedures.'" *Id.* at 46 (citing Ex. 1004, Figs. 6, 8; Ex. 2013 ¶¶ 100–102). In addition, SoundClear argues that, even assuming that Shimomura's Figures 7 and 9 "were part of the same process," a person of ordinary skill in the art would have had no reason to alter Shimomura's explicit teachings regarding these two figures because Shimomura teaches away from Amazon's proposed modifications (e.g., by limiting the speech volume of the robot). *Id.* at 47–50 (citing Ex. 1004 ¶¶ 72, 73; Ex. 2013 ¶¶ 102–105).

In its Reply, Amazon contends that SoundClear's arguments that a person of ordinary skill in the art would not have been motivated to adjust the distances associated with the volume change table illustrated in Shimomura's Figure 9 to overlap with the distances associated with character string transformation determination table illustrated in Shimomura's Figure 7 are incorrect for two reasons. Pet. Reply 20. First, despite SoundClear's arguments to the contrary, Amazon asserts that Shimomura teaches or suggests that the process of determining the character string transformation illustrated in Shimomura's Figure 7 is the input to the process of determining volume change output illustrated in Shimomura's Figure 9 and, therefore, "they are used in conjunction." *Id.* at 20–21 (citing Ex. 1004 ¶¶ 59–70, Figs. 7, 8, 9, 13; Pet. 43–47; Ex. 1202 ¶ 27). Second, Amazon asserts that SoundClear's teaching away arguments include the following flaws: (i) those arguments are legally erroneous because, even though Shimomura may teach accomplishing something one way, this would not discourage a person of ordinary skill in the art from accomplishing it

another way; (ii) those arguments rest on faulty, unsupported assumptions, such as Shimomura's robot's volume level reaches a maximum limit at a distance of 250 cm away from conversation partner user; and (iii) those arguments do not address specifically the motivations to modify Shimomura that Amazon proposes in the Petition. *Id.* at 21–22 (citing PO Resp. 49–51; Ex. 1004 ¶¶ 57, 72; Pet. 48–50; Ex. 1002 ¶¶ 164–169).

In its Sur-reply, SoundClear reiterates its argument that the change in output illustrated in Shimomura's Figure 7 is not related to the change in volume illustrated in Shimomura's Figure 9. PO Sur-reply 19–20 (citing Ex. 1004, Figs. 7, 9; PO Resp. 45–46). SoundClear contends that Amazon's proposed modification to Shimomura and, in particular, the distances associated with the volume change table illustrated in Shimomura's Figure 9, is based on impermissible hindsight reconstruction because "Shimomura expressly teaches . . . further increase in volume is not possible." *Id.* at 20–21 (citing Ex. 1004 ¶¶ 72, 73; PO Resp. 44–51). SoundClear asserts that "[t]he mere fact that Shimomura does not explicitly disparage setting the same distance for changing output volume and transforming output content" does not change SoundClear's initial position that Amazon's proposed modification "would directly undermine the embodiments [Shimomura] expressly discloses." *Id.* at 21.

Based on the complete record, we agree with Amazon that it would have been obvious to a person of ordinary skill in the art to adjust the distances associated with the volume change table illustrated in Shimomura's Figure 9 so that they overlap with the distances associated with character string transformation determination table illustrated in Shimomura's Figure 7. *See* Pet. 46–53; Pet. Reply 20–22. As an initial

matter, Amazon provides sufficient reasons why a person of ordinary skill in the art would have been motivated to modify Shimomura so that the distance thresholds in Shimomura's Figures 7 and 9 better align, which we adopt as our own. *See* Pet. 47–50. As one example, Amazon argues—and we agree—that such a modification would have been a simple substitution of a known element (i.e., a 350 cm distance threshold from the character string transformation determination table illustrated in Shimomura's Figure 7) for another known element (i.e., a 250 cm distance threshold from the volume change table illustrated in Shimomura's Figure 9) to obtain predictable results (i.e., a robot that adjusts speech volume based on whether the distance between a conversation partner and the robot is less than or greater than 350 cm). *See* Pet. 48 (citing Ex. 1002 ¶ 164).

SoundClear's arguments that Shimomura's Figures 7 and 9 are incompatible turns on whether there are certain limits to the speech volume of the robot. Shimomura discloses providing speech control for the robot by transforming the character string of output content so that it “can be clearly understood according to the physical distance to the conversation partner.” Ex. 1004 ¶ 57. Later, Shimomura discloses that “there is a certain limit to the speech volume of the robot” because “the content of the utterance of robot 1 may be difficult to hear.” *Id.* ¶ 72.

Shimomura, however, does not identify the specific limits of the speech volume of the robot that would impair the ability for the conversation partner to hear the output content provided by the robot. Indeed, we find it difficult to comprehend the implication drawn from SoundClear's argument that the robot reaches a maximum speech volume, such that any increase in speech volume is not possible beyond the user-to-device distance threshold

of 250 cm as illustrated in Figure 9. *See* PO Resp. 47; PO Sur-reply 20. It is difficult to reconcile SoundClear’s argument in this regard with Shimomura’s Figure 7 and its corresponding description because this figure indicates that a character string transformation occurs when the distance between a conversation partner user and the robot is more than 350 cm. In our view, if the distance between the user and the robot is more than 350 cm, the ability for the robot to respond accordingly would be severely hindered, if not impossible, if the robot were unable to increase the speech volume of the output content provided back to the conversation partner when, as SoundClear asserts, the user-to-device distance threshold exceeds more than 250 cm.

We also do not agree with SoundClear’s teaching away arguments. *See* PO Resp. 49–51. It is well-settled that a reference teaches away from a claimed invention if it “criticize[s], discredit[s], or otherwise discourage[s]” modifying the reference to arrive at the claimed invention. *In re Fulton*, 391 F.3d 1195, 1201 (Fed. Cir. 2004). We will not, however, “read into a reference a teaching away from a process where no such language exists.” *DyStar Textilfarben GmbH & Co. Deutschland KG v. C.H. Patrick Co.*, 464 F.3d 1356, 1364 (Fed. Cir. 2006). SoundClear’s teaching away argument is, once again, predicated on the notion that the robot reaches a maximum speech volume, such that any increase in speech volume is not possible beyond the user-to-device distance threshold of 250 cm illustrated in Figure 9. But as we explain above, if Shimomura’s robot employs software for analyzing the speech of a conversation partner user received as voice input from microphones 21 when the distance between the user and the robot is more than 350 cm, then the robot appears capable of increasing the speech

volume of the output content provided back to the user at the same distance of 350 cm. In our view, SoundClear attempts to read into Shimomura a teaching away where no such language exists. *See Dystar*, 464 F.3d at 1364.

d. Summary

In summary, Amazon has demonstrated by a preponderance of the evidence that the subject matter of independent claims 1, 4, and 5 would have been obvious over the teachings of Shimomura.

4. Claims 2 and 3

Claim 2 depends from independent claim 1, and further recites the following:

- 2[a] the process executing unit comprises:
 - an intention analyzing unit configured to extract intention information indicating an intention of the user based on the voice acquired by the voice acquiring unit; and
 - an acquisition content information acquiring unit configured to acquire acquisition content information which is notified to the user based on the extracted intention information, and
- 2[b] the voice-content generating unit is further configured to generate the text data including the acquisition content information as the output sentence.

Ex. 1001, 19:4–17 (with bracket annotations provided by Amazon in the Petition). Relying on the same arguments and evidence it presents for limitation 1[e], Amazon contends that Shimomura teaches or suggests limitation 2[a] because control unit 10 determines the output content by extracting and responding to a user’s intention. Pet. 53. More specifically, Amazon argues that Shimomura’s robot controls its speech “during a dialogue between the device and a user,” thereby enabling the robot to “converse with the user smoothly.” *Id.* (first quoting Ex. 1004 ¶ 10, then

quoting Ex. 1004 ¶ 86). As one example, Amazon asserts that Shimomura's control unit 10 acquires content such as "Today's weather in Tokyo" and presents this content to the user. *Id.* (quoting Ex. 1004 ¶ 65) (citing Ex. 1004 ¶ 44; Ex. 1002 ¶ 176).

Alternatively, Amazon contends that, even if Shimomura does not disclose limitations 2[a] and 2[b] either explicitly or inherently, it would have been obvious to a person of ordinary skill in the art to include an "intention analyzing unit" and an "acquisition content information acquiring unit" in Shimomura because it discloses "voice processing capabilities that allow the device to interpret user intent and generate appropriate responses." Pet. 54–55 (citing Ex. 1002 ¶ 173). Relying on the testimony of its declarant, Dr. Stern, Amazon argues that a person of ordinary skill in the art would have been motivated to incorporate an "intention analyzing unit" and an "acquisition content information acquiring unit" into Shimomura's control unit 10 because doing so would allow it "to perform these functions more efficiently," this "modular design would facilitate consistent intent extraction, reliable content retrieval, and improved clarity and organization of internal processes," and this would "merely reflect the way known voice analysis systems worked at the time of the invention." *Id.* at 55 (citing Ex. 1002 ¶ 180). Amazon asserts that modifying Shimomura's control unit 10 in this way "would have been straightforward" to a person of ordinary skill in the art because it "would have been an obvious design choice" that "enhanc[es] modularity, maintainability, and troubleshooting" to "improve response accuracy, while offering greater flexibility to handle various user response." *Id.* (citing Ex. 1002 ¶ 181).

Amazon contends that Shimomura teaches or suggests limitation 2[b] because control unit 10 acquires information, such as weather information, that is incorporated into the robot's voice content provided to the user. Pet. 56 (citing Ex. 1004 ¶¶ 10, 65, 86). More specifically, Amazon argues that Shimomura's output management unit 46 "sends the character string provided by the behavior selection control unit 45 to the speech synthesis unit 48, while the speech synthesis unit 48 generates a speech signal S4 of synthetic speech based on the supplied character string and sends it to the speaker[s] 23" for output. *Id.* (quoting Ex. 1004 ¶ 47).

Claim 3 depends from independent claim 1, and further recites "wherein, on generating the second sentence, the voice-content generating unit is further configured to omit a part of information included in the voice spoken by the user." Ex. 1001, 19:18–21. Relying on the same arguments and evidence it presents for limitations 1[e] and 1[g], Amazon contends that Shimomura teaches or suggests this limitation because control unit 10 generates content while conversing with the user (e.g., "Today's weather in Tokyo is sunny"). Pet. 56. Relying on the testimony of its declarant, Dr. Stern, Amazon argues that a person of ordinary skill in the art would have understood that generating this content would have been prompted by the user's speech (e.g., "what is the weather today"). *Id.* (citing Ex. 1002 ¶ 187). According to Amazon, "the interrogative word prompts processing, and the device omits it in response to provide a direct answer rather than echoing the question." *Id.*

During trial, SoundClear does not address separately Amazon's arguments and supporting evidence explaining how the teachings of Shimomura account for the additional limitations of dependent claims 2 and

3. *See* PO Resp. 37–52. We have reviewed Amazon’s arguments and supporting evidence explaining how the teachings of Shimomura account for the additional limitations of these dependent claims, and we agree with and adopt Amazon’s analysis. *See* Pet. 53–58. Accordingly, SoundClear has demonstrated by a preponderance of the evidence that the subject matter of dependent claims 2 and 3 would have been obvious over the teachings of Shimomura.

D. Obviousness over the Combined Teachings of Shimomura and Shin

Amazon contends that claims 1–5 of the ’337 patent are unpatentable under 35 U.S.C. § 103 as obvious over the combined teachings of Shimomura and Shin. Pet. 59–64. Amazon contends that the combined teachings of Shimomura and Shin account for the subject matter of each challenged claim, and Amazon provides reasoning as to why a person of ordinary skill in the art would have been prompted to modify the teachings of these references. *Id.* To support its positions, Amazon submits a Declaration of Dr. Stern with the Petition and a Reply Declaration of Dr. Stern with the Reply. *See* Exs. 1002, 1202.

Relying on the same arguments it presents to rebut the anticipation ground based on Shin, the obviousness ground based on Shin alone, and the obviousness ground based on Shimomura alone, SoundClear asserts that the combined teachings of Shimomura and Shin do not teach or suggest limitations 1[b], 1[f], 1[g], and 1[h]. PO Resp. 52. SoundClear also argues that Amazon does not provide a sufficient rationale to combine the teachings of Shimomura with those of Shin. PO Resp. 53–54. To support its positions, SoundClear submits the Declaration of Mr. McAlexander with the Patent Owner Response. *See* Ex. 2013.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1–5 of the ’337 patent are unpatentable as obvious over the combined teachings of Shimomura and Shin. We begin by addressing to the parties’ contentions with respect to the challenged claims.

1. Claims 1, 4, and 5

Once again, our analysis below focuses on the limitations of independent claim 1 by relying on the bracketed annotations provided by Amazon in the Petition. *See* Pet. ii, 59–62. For the most part, the same arguments and evidence Amazon provides to explain how the combined teachings of Shimomura and Shin account for all the limitations of independent claim 1 equally apply to all the corresponding limitations of independent claims 4 and 5. *See* Pet. 64. In addition, with the exception of limitation 1[h], Amazon’s arguments and evidence it identifies above with respect to its obviousness ground based on the teachings of Shimomura alone equally apply to its obviousness ground challenging these same claims based on the combined teachings of Shimomura and Shin. *See id.* at 59 (arguing “Shimomura . . . individually disclose[s] and render[s] obvious all limitations of claim 1”).

a. Limitations 1[b], 1[f], and 1[g]

As an initial matter, SoundClear asserts that the combined teachings of Shimomura and Shin do not teach or suggest limitations 1[b], 1[f], and 1[g]. PO Resp. 52. SoundClear, however, relies, at least in part, on the same arguments it presents to rebut the obviousness ground based on Shimomura alone. For the same reasons set forth above in our analysis of

the asserted grounds based on obviousness over Shimomura alone, we do not agree with SoundClear's arguments. *See supra* Section II.C.3.a and b.

b. Limitation 1[h]

With respect to limitation 1[h], Amazon contends that, “[e]ven if Shimomura alone [does] not teach or motivate a [person of ordinary skill in the art] to alter the output’s content and volume at the same threshold distance (e.g., 350 cm), doing so would have been obvious in view of Shin.” Pet. 59 (citing Ex. 1002 ¶¶ 199–204). Amazon argues that a person of ordinary skill in the art would have combined the teachings of Shimomura and Shin “to enhance adaptability and functionality in voice-controlled devices” because they both “teach proximity-based voice output adjustments for improving user interaction.” *Id.* at 60 (citing Ex. 1002 ¶ 200). Amazon further argues that Shin’s Table 1 and Table 3 both “include matching distance ranges, correlating specific distances with volume levels and information [output] amounts.” *Id.* (citing Ex. 1002 ¶ 201; Ex. 1003, Tables 1, 3). Amazon asserts that a person of ordinary skill in the art would have been motivated to modify Shimomura’s volume change table illustrated in Figure 9 “to implement common distance thresholds for adjusting volume level and [output] content.” *Id.* at 61 (citing Ex. 1002 ¶ 201). According to Amazon, “[t]his would enable Shimomura’s device to classify a voice into one of two distance-based categories and adjust output volume level and content based on that category, rather than referencing tables with distance ranges.” *Id.*

In its Patent Owner Response, SoundClear contends that neither Shimomura nor Shin teach or suggest the volume adjustment required by limitation 1[h] because “nothing in Shin would have motivated a [person of ordinary skill in the art] to deviate from Shimomura’s express[] teachings.” PO Resp. 52–53. At best, SoundClear argues “Shin suggests to a [person of ordinary skill in the art] that the user-to-device distance at which to change output volume and output content *could* be the same,” but a person of ordinary skill in the art would not have applied this teaching to Shimomura because “doing so would have been undesirable as it would have burdened Shimomura’s robot and user with superfluous phrases (e.g., ‘ne’) when they are unnecessary for the robot to be understood.” *Id.* at 53 (citing Ex. 2013 ¶ 111). SoundClear also contends that Amazon engages in impermissible hindsight reconstruction because “Shin, like Shimomura, discloses tables with more than two distance ranges.” *Id.* (citing Ex. 2013 ¶ 112). SoundClear asserts that “[n]either Shin nor Shimomura discloses classifying a voice into one of two categories.” *Id.*

In its Reply, Amazon contends that SoundClear admits that Shimomura and Shin could be combined, but suggests there is no reason to do so, effectively ignoring Amazon’s rationales to combine the teachings of Shimomura and Shin proposed in the Petition. Pet. Reply 23–24 (citing PO Resp. 53; Pet. 59–64). As one example, Amazon argues that a person of ordinary skill in the art would have been motivated to modify Shimomura to include Shin’s classification scheme that includes the same distance ranges so that Shimomura’s classification scheme “likewise occur[s] at the same distance ranges.” *Id.* at 24 (citing Pet. 60–61). Amazon also disagrees with SoundClear’s argument that neither Shin nor Shimomura disclose

classifying a voice into one of two categories because the character string transformation determination table in Shimomura's Figure 7 clearly shows only two categories. *Id.* (citing Ex. 1004, Fig. 7; Ex. 1201, 112:13–113:2).

In its Sur-reply, SoundClear maintains that “[n]either Shimomura nor Shin disclose[] the claimed volume adjustment” required by limitation 1[h]. PO Sur-reply 21. SoundClear contends that a person of ordinary skill in the art would not have been motivated “to modify Shimomura to set the distance at which to change speech volume to be the same as the distance at which to transform the output content” because, at best, Shin suggests that “user-to-device distance at which to change output volume and output content *could* be the same,” but this suggestion would be “one of an almost unlimited number of options.” *Id.* at 22.

Based on the complete record, we agree with Amazon that, even if Shimomura alone does not teach or suggest the same distance thresholds (e.g., 350 cm) for adjusting the output content and volume level, Shin nevertheless teaches the same distance thresholds (e.g., every 1 m) for adjusting the output content and volume level and, therefore, the combined teachings of Shimomura and Shin properly account for limitation 1[h]. In our analysis of the anticipation ground based on Shin and the obviousness ground based on Shin alone, we find that Shin's Table 3 sets forth several levels of classification based on distance thresholds, one of which classifies the user's voice as a first voice when the distance between the user and electronic device 101 is less than 1 meter, and another that classifies the user's voice as a second voice when the distance between the user and electronic device 101 is more than 4 meters. *See supra* Section II.B.4.a.i. We also find that, depending on the distance between a user providing voice

input and electronic device 101, Shin discloses increasing the output volume level by 5 dB for every 1 meter the user moves away from electronic device 101. *See supra* Section II.B.4.c.

Together, these findings demonstrate the same distance thresholds (i.e., every 1 meter) between the output volume level of content illustrated in Shin’s Table 1 and the output amount of information of content illustrated in Shin’s Table 3. Indeed, SoundClear appears to admit that Shin teaches using the same distance thresholds for adjusting the output content and volume level. *See, e.g.*, PO Resp. 53 (“At best, Shin suggests to a [person of ordinary skill in the art] that the user-to-device distance at which to change the output volume and output content *could* be the same.”); PO Sur-reply 22 (“[A]t best [Shin] suggests that the user-to-device distance at which to change the output volume and output content *could* be the same.”).

With this in mind, Amazon argues—and we agree—that it would have been obvious to a person of ordinary skill in the art to modify Shimomura’s volume change table illustrated in Figure 9 to implement the same distance thresholds for adjusting volume level as those applied to the output content in Shimomura’s character string transformation determination table illustrated in Figure 7. *See* Pet. 61 (citing Ex. 1002 ¶ 201). A person of ordinary skill in the art would have been motivated to modify Shimomura in this way because Shin provides a teaching, suggestion, or motivation for doing so—namely, Shin teaches using the same distance thresholds for adjusting the output content and volume level. *Compare* Ex. 1004 ¶ 86, Table 3, *with* Ex. 1004 ¶ 79, Table 1.

We do not agree with SoundClear’s argument that Amazon engages in impermissible hindsight reconstruction. *See* PO Resp. 53–54. SoundClear

does not direct us to persuasive evidence showing that actually applying Shin's teaching of using the same distance thresholds for adjusting the output content and volume level to the tables illustrated in Shimomura's Figures 7 and 9 would have been "uniquely challenging or difficult for one of ordinary skill in the art," which is a beneficial check on hindsight.

Leapfrog Enters., Inc. v. Fisher-Price, Inc., 485 F.3d 1157, 1162 (Fed. Cir. 2007) (citing *KSR*, 550 U.S. at 419).

We also do not agree with SoundClear's argument that a person of ordinary skill in the art would not have been motivated to modify Shimomura in manner proposed by Amazon because neither Shin nor Shimomura discloses classifying a voice into one of two categories. *See* PO Resp. 53–54. Amazon argues—and we agree—that Shimomura's character string transformation determination table illustrated in Figure 7 clearly illustrates classifying a voice into one of two categories. *See* Pet. Reply 24. That is, we find that Shimomura's Figure 7 sets forth two levels of classification based on distance thresholds, one of which classifies the user's voice as a first voice when the distance between the user and the robot is more than 350 cm, and another that classifies the user's voice as a second voice when the distance between the user and the robot is less than 350 cm. *See supra* Section II.C.3.A.i.

c. Summary

In summary, Amazon has demonstrated by a preponderance of the evidence that independent claims 1, 4, and 5 would have been obvious over the combined teachings of Shimomura and Shin.

2. Claims 2 and 3

During trial, SoundClear does not address separately Amazon's arguments and supporting evidence explaining how the combined teachings of Shimomura and Shin account for the additional limitations of dependent claims 2 and 3. *See* PO Resp. 52–54. We have reviewed Amazon's arguments and supporting evidence explaining how the combined teachings of Shimomura and Shin account for the additional limitations of these dependent claims, and we agree with and adopt Amazon's analysis. *See* Pet. 62–64. Accordingly, Amazon has demonstrated by a preponderance of the evidence that the subject matter of dependent claims 2 and 3 would have been obvious over the combined teachings of Shimomura and Shin.

E. Obviousness over the Combined Teachings of Shin and/or Shimomura in View of Kristjansson

Amazon contends that claims 1–5 of the '337 patent are unpatentable under 35 U.S.C. § 103 as obvious over the combined teachings of Shin and/or Shimomura in view of Kristjansson. Pet. 65–68. Amazon contends that the combined teachings of Shin and/or Shimomura in view of Kristjansson account for the subject matter of each challenged claim, and Amazon provides reasoning as to why a person of ordinary skill in the art would have been prompted to modify the teachings of these references. *Id.* To support its positions, Amazon submits a Declaration of Dr. Stern with the Petition and a Reply Declaration of Dr. Stern with the Reply. *See* Exs. 1002, 1202.

Relying on the same arguments it presents to rebut the other grounds of unpatentability raised by Amazon in the Petition, SoundClear asserts that the combined teachings of Shin and/or Shimomura in view of Kristjansson do not teach or suggest limitations 1[b], 1[f], 1[g], and 1[h]. PO Resp. 54.

To support its positions, SoundClear submits the Declaration of Mr. McAlexander with the Patent Owner Response. *See* Ex. 2013.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1–5 of the ’337 patent are unpatentable as obvious over the combined teachings of Shin and/or Shimomura in view of Kristjansson. We begin by providing an overview of Kristjansson, and then we turn to the parties’ contentions with respect to the challenged claims.

1. Overview of Kristjansson

Kristjansson generally relates to speech processing systems that combine speech recognition and natural language understanding (“NLU”) processing techniques. Ex. 1013, 1:6–22. Kristjansson discloses performance drawbacks associated with speech processing due to “listening environments for users” that “may vary in terms of the amount of noise experienced by the user” that “make it difficult for a user to hear and/or understand output audio from a device.” *Id.* at 2:27–30. To address these drawbacks, Kristjansson discloses “a system [that] may determine a location of a user and noise sources relative to a device, may determine the noise level at the user’s location, and may determine a gain for output audio data to ensure the output audio is at a sufficient volume when it reaches the user’s location.” *Id.* at 2:30–36.

Kristjansson discloses a user in an environment with device 110 and noise sources 190a/190b, where “device 110 may detect (130) the input audio which may include both the utterance as well as noise from the noise sources” and “may send (132) audio data to the server 120 for speech processing.” Ex. 1013, 3:1–8, Fig. 1. Server 120 may include, among other

things, NLU processing 460 with “various components, including potentially dedicated processor(s), memory, storage, etc.” *Id.* at 7:26–28, Fig. 4. “[T]he NLU process takes textual input (such as processed from ASR [(“automatic speech recognition”)] 450 based on the utterance input audio 11 [from the user]) and . . . derive[s] an intent or a desired action from the user.” *Id.* at 7:49–63.

2. Claims 1–5

Once again, our analysis below relies on the bracketed annotations provided by Amazon in the Petition. Relying on the arguments and evidence presented for the obviousness grounds based on Shin and/or Shimomura, Amazon contends that the Shin, Shimomura, or both references teach or suggest each limitation of claims 1–5. Pet. 65. Amazon argues that, to the extent that Shin, Shimomura, or both references do not teach or suggest the claimed “processing executing unit” required by limitations 1[c], 2[a], 4[d], and 5[d], Kristjansson teaches or suggests those limitations. *Id.* (citing Ex. 1002 ¶¶ 216–223). As one example, Amazon directs us to limitation 2[a], which recites, in relevant part, the “process executing unit” comprises an “intention analyzing unit” and an “acquisition content information acquiring unit,” arguing that these limitations determine the user’s intent using NLU processing and acquire the desired information. *Id.* (citing Ex. 1001, 4:31–33; Ex. 1002 ¶ 217). Amazon contends that Kristjansson teaches a speech processing system similar to those taught by Shin and Shimomura, but it includes “an NLU component” that “takes textual input (such as processed from ASR 450 based on the utterance input audio 11 [from the user]) and . . . derive[s] an intent or a desired action from the user.” *Id.* at 65–67 (quoting Ex. 1013, 7:26–67). As a result, Amazon

asserts that Kristjansson teaches or suggests limitation 2[a], as well as the broader limitations of the other challenged claims, including limitations 1[c], 4[d], and 5[d]. *Id.* at 66.

Turning to rationale to combine and reasonable expectation of success, Amazon contends that a person of ordinary skill in the art would have been motivated to implement Kristjansson’s NLU teachings in the devices of Shin and/or Shimomura to process the user’s input and determine/retrieve the proper output for several reasons. Pet. 66 (citing Ex. 1002 ¶¶ 219–222). As one example, Amazon argues that a person of ordinary skill in the art would have looked to Kristjansson to implement NLU teachings in Shin and/or Shimomura because “Shin and Shimomura both disclose systems that analyze voice input to identify the content the user is interested in, and then generate the requested content.” *Id.* According to Amazon, Kristjansson “merely provides additional detail about how a [person of ordinary skill in the art] would structure such a system, with NLU processing and content retrieval.” *Id.* at 66–67 (citing Ex. 1002 ¶ 220). Amazon asserts that a person of ordinary skill in the art would have had a reasonable expectation of success in implementing Kristjansson’s NLU teachings in Shin and/or Shimomura because “Kristjansson’s approach was well-known in the art,” “Shin and Shimomura contemplate modification or addition to their systems by [a person of ordinary skill in the art],” and all three references “use[] a proximity sensor to determine the user-to-device distance and adjust[] the volume of the output based on that distance.” *Id.* at 67 (citing Ex. 1015, 431–33; Ex. 1003 ¶ 184; Ex. 1004 ¶ 89; Ex. 1013, claim 1, 3:1–27; Ex. 1002 ¶ 222).

a. Limitations 1[b], 1[f], 1[g], and 1[h]

SoundClear asserts that the combined teachings of Shimomura and/or Shin in view of Kristjansson do not teach or suggest limitations 1[b], 1[f], 1[g], and 1[h]. PO Resp. 54–55. SoundClear, however, relies, on the same arguments it presents to rebut the other asserted grounds raised by Amazon in the Petition. For the same reasons set forth above, we do not agree with SoundClear’s arguments. *See supra* Section II.B.4 and 5; II.C.3 and 4; II.D.1 and 2.

b. Summary

In summary, Amazon has demonstrated by a preponderance of the evidence that claims 1–5 would have been obvious over the combined teachings of Shin and/or Shimomura in view of Kristjansson.

F. Remaining Arguments

In its Patent Owner Response, SoundClear presents additional arguments as to why Amazon’s obviousness analysis set forth in the Petition is legally deficient. PO Resp. 55-65. We address SoundClear’s arguments in turn.

1. Illogical Modifications

SoundClear contends that “Shin and Shimomura are concerned only with making a discrete measurement of distance between a device and user,” and given the simplicity of this metric, “Shin and Shimomura specifically declined to conduct an analysis of voices to classify voices, let alone based on distance.” PO Resp. 55 (citing Ex. 2013 ¶ 114). SoundClear argues that Shin and Shimomura each “purposely chose simpler, more discrete approaches” and Amazon’s proposed modifications to these two references “would have achieved the opposite.” *Id.* In its Reply, Amazon contends

that both Shin and Shimomura “disclose[] classifying voice input by distance in the same way, and their simplicity supports their combination.” Pet. Reply 25 (citing Ex. 1202 ¶ 29).

As an initial matter, SoundClear repeats essentially the same argument it presents for each of the obviousness grounds raised by Amazon in the Petition. For the same reasons we identify above in our analysis of these obviousness grounds, we do not agree with SoundClear’s argument. *See supra* Section II.B.4.a.i and ii; Section II.C.3.i and ii. In addition, we agree with Amazon that, because Shin and Shimomura both teach classifying voice input by distance in the same way (i.e., by using distance thresholds between a user and a device), this further supports Amazon’s position that it would have been obvious to a person of ordinary skill in the art to modify Shimomura’s volume change table illustrated in Figure 9 to implement the same distance thresholds for adjusting volume level as those applied to the output content in Shimomura’s character string transformation determination table illustrated in Figure 7. *See* Pet. 61 (citing Ex. 1002 ¶ 201). More specifically, a person of ordinary skill in the art would have been motivated to modify Shimomura in this way because Shin provides a teaching, suggestion, or motivation for doing so—namely, Shin teaches using the same distance thresholds for adjusting the output content and output volume level. *Compare* Ex. 1004 ¶ 86, Table 3, *with* Ex. 1004 ¶ 79, Table 1.

2. *Graham Factor Two*

SoundClear contends that “[Amazon] fail[s] to apply *Graham* factor two in [its] obviousness analysis” because Amazon does “not explain what is specifically missing from the primary references or why a [person of ordinary skill in the art] would have been motivated to bridge that particular

gap.” PO Resp. 56–57. According to SoundClear, this is evident from deposition testimony of Amazon’s declarant, Dr. Stern, because he was unable to articulate any difference between Shin or Shimomura and independent claim 1. *Id.* at 57 (citing Ex. 2015, 109:12–112:16, 113:3–117:5). In its Reply, Amazon contends that the Petition applies the *Graham* factors in its obviousness analysis, and “Dr. Stern’s [deposition] testimony does not alter this analysis” because, at least with respect to Shin, “there are no differences to articulate.” Pet. Reply 26.

Beginning with Shin, there are no difference between Shin and the challenged claims. *See supra* Section II.B.1–5. Dr. Stern’s deposition testimony appears to reinforce Amazon’s position that Shin anticipates and renders obvious the challenged claims because he testifies there is “no difference” between Shin and the challenged claims. *See* Ex. 2015, 109:12–112:16. Turning to Shimomura, the only difference between the teachings of Shimomura and the challenged claims applies to limitation 1[h] and, in particular, whether the distance thresholds associated with the volume change table illustrated in Shimomura’s Figure 9 overlap with the distance thresholds associated with character string transformation determination table illustrated in Shimomura’s Figure 7. Amazon, however, specifically identifies this difference in the Petition. *See* Pet. 47. Moreover, as we explain above, Amazon presents a sufficient rationale to modify Shimomura so that Shimomura’s volume change table illustrated in Figure 9 implements the same distance thresholds for adjusting volume level as those applied to the output content in Shimomura’s character string transformation determination table illustrated in Figure 7. *See supra* Section II.C.3.c.

Bottom line, we are convinced that Amazon correctly applies *Graham* factor two in its obviousness analysis.

3. *No Reason for the Combinations*

SoundClear contends that “[Amazon] indiscriminately select[s] disclosures from the various references to attempt to piece together the [challenged claims],” but Amazon never specifically addresses why a person of ordinary skill in the art “would have seen any reason to combine or modify the references to fill in those gaps.” PO Resp. 59. More specifically, when addressing Amazon’s proposed combination of Shimomura and Shin, SoundClear contends that Amazon focuses on the two tables illustrated in Shimomura’s Figures 7 and 9 “without any discussion of the larger purpose or content of any claimed invention.” *Id.*

We disagree with SoundClear’s argument that Amazon does not provides a sufficient rationale to combine the teachings of Shimomura and Shin. Once again, as we explain above, we agree with Amazon that it would have been obvious to a person of ordinary skill in the art to modify Shimomura’s volume change table illustrated in Figure 9 to implement the same distance thresholds for adjusting volume level as those applied to the output content in Shimomura’s character string transformation determination table illustrated in Figure 7. *See* Pet. 61 (citing Ex. 1002 ¶ 201). More specifically, a person of ordinary skill in the art would have been motivated to modify Shimomura in this way because Shin provides a teaching, suggestion, or motivation for doing so—namely, Shin teaches using the same distance thresholds for adjusting the output content and output volume level. *Compare* Ex. 1004 ¶ 86, Table 3, *with* Ex. 1004 ¶ 79, Table 1.

4. Perfunctory Obviousness Arguments

SoundClear contends that Amazon’s obviousness analysis “does not establish that any results would have been predictable [nor] any reasonable expectation of success” because its statements on these two issues “lack support and are legally insufficient.” PO Resp. 59–60. In its Reply, Amazon contends that it provides “many motivations to combine and modify references and explained why a [person of ordinary skill in the art] would have reasonably expected success.” Pet. Reply 25 (citing Pet. 48–50).

We do not agree with SoundClear that Amazon’s obviousness arguments were perfunctory, lacked support, or were not legally insufficient. As one example, relying on the testimony of its declarant, Dr. Stern, Amazon provides six different rationales to combine the teachings of Shimomura and Shin and explains why there was a reasonable expectation of success in doing so, none of which SoundClear squarely addresses in its Patent Owner Response or its Patent Owner Sur-reply. *See* Pet. 48–50; Ex. 1002 ¶¶ 164–170.

5. Whether the Federal Circuit has Rejected Amazon’s Proffered Motivations to Combine

SoundClear contends that Amazon’s rationales to combine “are purely generic and bear no relation to the specific claims” because “[t]hey lack any particular meaning in an obviousness analysis.” PO Resp. 60, 62. SoundClear then proceeds to identify certain statements pertaining to Amazon’s rationales to combine set forth in the Petition that it labels as “generic, and less specific,” arguing that the Federal Circuit has rejected using similar rationales to combine in other cases. *Id.* at 61–62. In its Reply, Amazon contends that SoundClear’s argument in this regard

effectively “strip[s] the reference-specific portions of the arguments away and then argu[es] that they are not reference-specific.” Pet. Reply 25.

We do not agree with SoundClear’s argument that Amazon’s rationales to combine are generic and bear no relation to the challenged claims. In our view, SoundClear merely provides snippets of Amazon’s rationale to combine arguments in isolation, but fails to appreciate the reference-specific nature of those arguments. *See* PO Resp. 61–62. This approach does not convince us that Amazon’s rationales to combine set forth in the Petition represent the type of generic statements that Federal Circuit has deemed insufficient to support a conclusion of obviousness.

6. Improper Circumvention of Obviousness and Anticipation

SoundClear contends that Amazon’s obviousness grounds based on the combined teachings of Shimomura and Shin, with or without Kristjansson, are presented in a “quick, cursory fashion,” such that they “appear to be attempts to avoid the proper obviousness and inherency standards.” PO Resp. 63. We, however, see no issue with Amazon presenting arguments and evidence for the obviousness grounds based on the combined teachings of Shimomura and Shin, with or without Kristjansson, by referring back to its prior arguments and evidence presented in the anticipation and obviousness grounds based on Shin, or the obviousness ground based on Shimomura alone. In general, this approach is permissible as long as it is clear.

7. Failure to Provide Fair Notice

SoundClear contends that both the Petition and Dr. Stern’s declaration accompanying the Petition are “confusing” and “non-committal” because they “repeatedly state contradictory alternative positions without

differentiating between the two: ‘known and obvious,’ ‘anticipate and render obvious,’ ‘disclose and render obvious,’ ‘disclose or render obvious.’” PO Resp. 63 (citing Pet. 59, 62, 64–65, 69; Ex. 1002 ¶¶ 175, 186, 198, 216, 217). In its Reply, Amazon contends that the statements identified by SoundClear are not contradictory because “a teaching may be both known and obvious.” Pet. Reply 26.

We do not agree with SoundClear’s argument that Amazon’s Petition and Dr. Stern’s declaration accompanying the Petition present confusing, non-committal positions that result in a lack of notice. Once again, SoundClear merely provides snippets of Amazon’s statements in isolation, but fails to appreciate those statements in their full context. *See* PO Resp. 63. This approach does not convince us that Amazon’s Petition fails to “identif[y], in writing and with particularity, each claim challenged [and] the grounds on which the challenge to each claim is based.” 35 U.S.C. § 312(a)(3).

8. Expert Declaration Inseparable from Attorney Argument

SoundClear contends that the testimony of Amazon’s declarant, Dr. Stern, should be given little to no weight because the language in the Petition mirrors statements in Dr. Stern’s declaration testimony. PO Resp. 64. We do not agree. Merely because Amazon’s Petition may mirror Dr. Stern’s declaration testimony in certain respects would not necessarily result in us discounting his declaration testimony in its entirety. *See, e.g., Hum Indus. Tech., Inc. v. Amsted Rail Co., Inc.*, IPR2023-00540, Paper 10 at 37 (PTAB Aug. 11, 2023) (“If declaration testimony is supported sufficiently and, in turn, supports contentions in the Petition, we do not give it little to no weight simply because it is repeated in the Petition.”).

Put simply, we are capable of weighing Dr. Stern’s testimony and crediting it where appropriate. *See* 37 C.F.R. § 42.65 (“Expert testimony that does not disclose the underlying facts or data on which the opinion is based is entitled to little or no weight.”).

9. Modifying Outputs Does Not Disclose Classification of Inputs

SoundClear contends that Shin’s and Shimomura’s disclosures of “[a]djusting the outputs does not address the missing limitation of classifying inputs (voices).” PO Resp. 65. This is essentially the same argument raised by SoundClear to rebut the anticipation and obviousness grounds based on Shin, and the obviousness ground based on Shimomura alone. For the same reasons we identify above, we do not agree with SoundClear’s argument. *See supra* Section II.B.4.a.i and ii; Section II.C.3.a.i and ii.

III. CONCLUSIONS⁵

Based on the fully developed trial record, Amazon demonstrates by a preponderance of the evidence that claims 1–5 of the ’337 patent are unpatentable. A summary of our conclusions is set forth in the table below.

⁵ Should SoundClear wish to pursue amendment of the challenged claims in a reissue or reexamination proceeding after the issuance of this decision, we draw SoundClear’s attention to the April 2019 Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding. *See* 84 Fed. Reg. 16,654 (Apr. 22, 2019). If SoundClear chooses to file a reissue application or a request for reexamination of the challenged patent, we remind SoundClear of its continuing obligation to notify the Board of any such related matters in updated mandatory notices. *See* 37 C.F.R. § 42.8(a)(3), (b)(2).

Claims	35 U.S.C. §	Reference(s)/ Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1–5	102(a)(1)/ 102(a)(2)	Shin	1–5	
1–5	103	Shin ⁶		
1–5	103	Shimomura	1–5	
1–5	103	Shimomura, Shin	1–5	
1–5	103	Shin, Shimomura, Kristjansson	1–5	
Overall Outcome			1–5	

IV. ORDER

Accordingly, it is

ORDERED that claims 1–5 of the '337 patent have been shown to be unpatentable; and

FURTHER ORDERED that, because this is a Final Written Decision, parties to this proceeding seeking judicial review of our decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

⁶ Because we determine that claims 1–5 are unpatentable as anticipated by Shin, we need not reach Amazon's contingent obviousness ground over Shin challenging the same claims.

IPR2025-00565
Patent 11,069,337 B2

For PETITIONER

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

For PATENT OWNER

Steven J. Reynolds
Kevin H. Sprenger
Donald L. Jackson
Stephen McBride
Chandran B. Iyer
Erin Hadi
Louay Meroueh
DAIGNAULT IYER LLP
sreynolds@daignaultiyer.com
ksprenger@diagnaultiyer.com
djackson@diagnaultiyer.com
smcbride@diagnaultiyer.com
cbiyer@diagnaultiyer.com
ehadi@diagnaultiyer.com
lmeroueh@diagnaultiyer.com