

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.

IPR2025-00673
Patent 11,244,675 B2

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

Opinion for the Board filed by *Administrative Patent Judge* ZECHER.

Dissenting opinion filed by *Administrative Patent Judge* McKONE.

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 1–7 of U.S. Patent No. 11,244,675 B2 (Ex. 1001, “the ’675 patent”). Paper 1 (“Pet.”). Patent Owner, SoundClear Technologies LLC (“SoundClear”), filed a Preliminary Response. Paper 8 (“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 1–7 of the ’675 patent. Pursuant to 35 U.S.C. § 314, the Board instituted this IPR on October 2, 2025, as to claims 1–7 of the ’675 patent based on all grounds raised in the Petition. Paper 15 (“Dec. on Inst.”).

During trial, SoundClear filed a Patent Owner Response (Paper 20, “PO Resp.”),¹ Amazon filed a Reply (Paper 25, “Pet. Reply”), and SoundClear filed a Sur-reply (Paper 32, “PO Sur-reply”). An oral argument was held on July 8, 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 SoundClear also filed a Contingent Motion to Amend that proposed to replace originally challenged claim 5 of the ’675 patent with proposed substitute claim 8. Paper 21. We, however, granted SoundClear’s request to withdraw its Contingent Motion to Amend and, as a result, the Contingent Motion to Amend will no longer be considered in this proceeding. Paper 30.

1–7 of the ’675 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. 73; Paper 13 (Amazon’s Updated Mandatory Notices), 1. SoundClear identifies itself, which “is a wholly owned subsidiary of SoundClear LLC,” as an RPI. Paper 31 (SoundClear’s Updated Mandatory Notices), 1.

B. Related Matters

The parties indicate that the ’675 patent is the subject of, or the parties are involved in, the following three district court cases: (1) *SoundClear Techs. LLC v. Amazon.com Inc.*, No. 1:24-cv-01283 (E.D. Va. filed July 25, 2024); (2) *SoundClear Techs. LLC v. Google LLC*, No. 3:24-cv-00540 (E.D. Va. filed July 25, 2024); and (3) *SoundClear Techs. LLC v. Amazon.com Inc.*, 2:24-cv-00320 (E.D. Va. filed May 1, 2024). Paper 13 at 1–2; Paper 31 at 2. In addition to this Petition, Google LLC filed the following two other petitions challenging the ’675 patent: (1) *Google LLC v. SoundClear Techs. LLC*, IPR2025-01177, Paper 1 (PTAB July 14, 2025) (challenging claims 1–7 of the ’675 patent); and (2) *Google LLC v. SoundClear Techs. LLC*, IPR2026-00073, Paper 1 (PTAB Nov. 3, 2025) (challenging claims 1–7 of the ’675 patent). Paper 13 at 2.

Amazon also filed the following five petitions challenging related patents owned by SoundClear: (1) *Amazon.com, Inc. v. SoundClear Techs. LLC*, IPR2025-00565, Paper 1 (PTAB Jan. 31, 2025) (challenging claims 1–5 of U.S. Patent No. 11,069,337 B2); (2) *Amazon.com, Inc. v. SoundClear*

IPR2025-00673

Patent 11,244,675 B2

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 13 at 2.

C. The '675 Patent

The '675 patent, titled “Word Replacement in Output Generation for Detected Intent by Voice Classification,” issued from U.S. Patent Application No. 16/295,034, filed on March 7, 2019. Ex. 1001, codes (54), (21), (22). The '675 patent claims foreign priority to Japanese Patent Application No. JP2018-044598, filed on March 12, 2018. *Id.* at code (30).

The '675 patent generally relates to “an output-content control device, an output-content control method, and a non-transitory storage medium” that stores a program that causes a computer to execute certain steps. Ex. 1001, 1:31–34. The '675 patent acknowledges that the prior art discloses “voice control devices that analyze [the] detected voice of a user and perform . . . processing according to an intent of the user.” *Id.* at 1:20–24. The prior art also discloses “voice control devices that output [a] notification indicating that a processing intended by a user has been performed in voice, or that output an inquiry from a user in voice.” *Id.* at 1:24–27. The '675 patent, however, states it would be desirable for “a voice processing device that outputs voice” to provide a “notification in response to an inquiry from the

user [that is] difficult to be understood by people other than the user when the content of notification is output.” *Id.* at 1:28–37.

The '675 patent addresses this need by providing an output-content control device, an output content control method, and a non-transitory storage medium that stores a program that causes a computer to execute certain steps, all of which are capable of acquiring a voice, classifying the voice, analyzing the voice to detect intent information, acquiring notification information based on the intent information, and generating output based on the notification information. *See, e.g.*, Ex. 1001, 1:44–61 (disclosing “an output-content control device”), 1:62–2:8 (disclosing “an output-content control method”), 2:9–29 (disclosing “a non-transitory storage medium” that stores a program that causes a computer to execute certain steps). Figure 1, reproduced below, illustrates an output-content control device according to one embodiment. *Id.* at 2:33–34, 2:64–66.

FIG.1

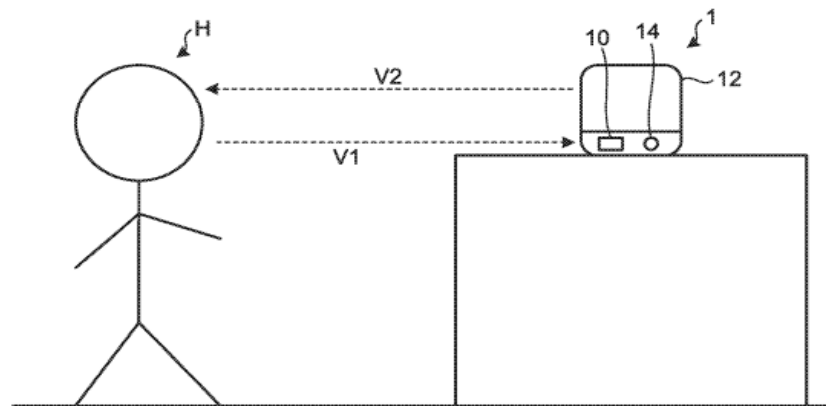


Figure 1 illustrates output-content control device 1 that detects voice V1 spoken by user H using voice detecting unit 10, analyzes detected voice V1

to perform pre-determined processing, and outputs voice V2 using voice output unit 12. *Id.* at 2:66–3:4.

Figure 2, reproduced below, illustrates a schematic block diagram of the output-content control device. Ex. 1001, 2:35–36, 3:17–18.

FIG.2

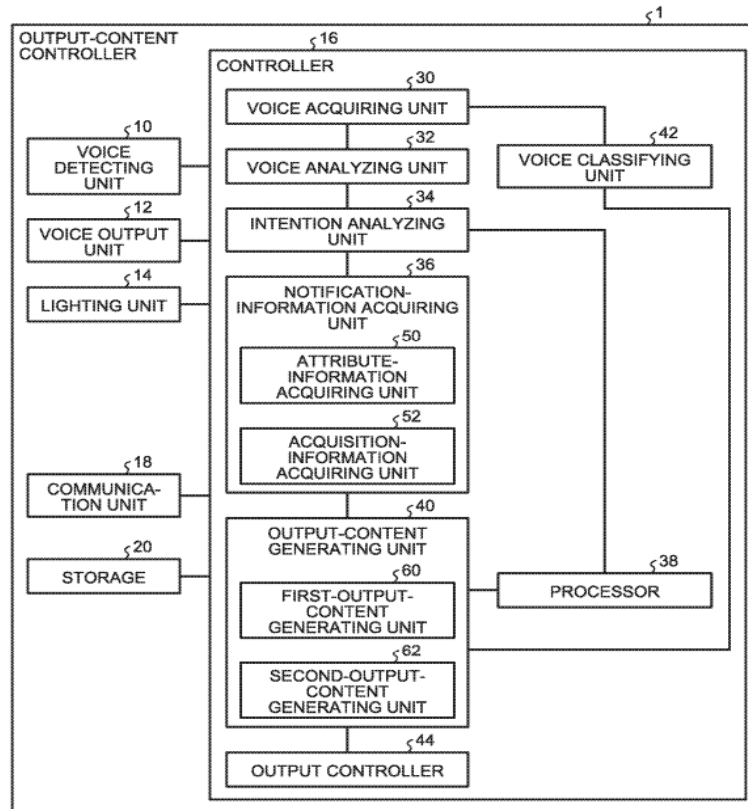


Figure 2 illustrates output-content control device 1 that includes voice detecting unit 10, voice output unit 12, lighting unit 14, controller 16, communication unit 18, and storage 20. *Id.* at 3:17–22. Controller 16 operates as a central processor unit (“CPU”) because it includes, among other things, voice acquiring unit 30, voice analyzing unit 32, intention analyzing unit 34, notification-information acquiring unit 36, processor 38, output-content generating unit 40, voice classifying unit 42, and output controller 44. *Id.* at 3:55–67.

D. Challenged Claims

Claims 1, 6, and 7 are independent. Independent claim 1 recites “[a]n output-content control device,” independent claim 6 recites “[a]n output-content control method,” and independent claim 7 recites “[a] non-transitory storage medium that stores an output-content control program that causes a computer to execute” certain steps. Ex. 1001, 20:60, 22:8–9, 22:29–30. Claims 2–5 directly or indirectly depend from independent claim 1. *Id.* at 21:21–22:7. Independent claim 1 is illustrative of the challenged claims and is reproduced below.

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

when the voice is determined to be the second voice,² a second output sentence which includes all of the intention information and the content information.

Id. at 20:60–21:20 (with bracketed annotations provided by Amazon in the Petition).

E. Asserted Prior Art References

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

Name³	Reference	Dates	Exhibit No.
Shin	US 2017/0083281 A1	published Mar. 23, 2017; filed Sept. 19, 2016	1003
Shimomura	JP 2005-202076 A	published Oct. 28, 2005; filed Jan. 14, 2004	1004 (English Translation)
Aoyama	US 2017/0337921 A1	published Nov. 23, 2017; Patent Cooperation Treaty (“PCT”) filed Nov. 26, 2015	1010
Kristjansson	US 10,147,439 B1	issued Dec. 4, 2018; filed Mar. 30, 2017	1013
Schuster	US 9,680,983 B1	issued June 13, 2017; filed June 16, 2016	1016
Iwase	US 11,183,167 B2	issued Nov. 23, 2021; PCT filed Dec. 25, 2017	1017

F. Asserted Grounds of Unpatentability

Amazon challenges claims 1–7 of the ’675 patent based on the asserted grounds of unpatentability set forth in the table below. Pet. 5, 7–69.

² See Certificate of Correction issued May 21, 2024 (changing “when the voice is not determined to be the second voice” to “when the voice is determined to be the second voice”).

³ For clarity and ease of reference, we only list the first named inventor.

Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
1–7	103 ⁴	Shin, Aoyama
1, 2, 5–7	103	Shin, Iwase
1–7	103	Shimomura, Aoyama
1, 2, 5–7	103	Shimomura, Iwase
3, 4	103	Shin, Iwase, Aoyama
3, 4	103	Shimomura, Iwase, Aoyama
5	103	Shin, Aoyama, Schuster
5	103	Shin, Iwase, Schuster
5	103	Shimomura, Aoyama, Schuster
5	103	Shimomura, Iwase, Schuster
1–7	103	Shin, Aoyama, Kristjansson
1–7	103	Shin, Iwase, Kristjansson
1–7	103	Shimomura, Aoyama, Kristjansson
1–7	103	Shimomura, Iwase, Kristjansson
1–7	103	Shin, Iwase, Aoyama, Kristjansson
1–7	103	Shimomura, Iwase, Aoyama, Kristjansson
1–7	103	Shin, Aoyama, Schuster, Kristjansson
1–7	103	Shin, Iwase, Schuster, Kristjansson
1–7	103	Shimomura, Aoyama, Schuster, Kristjansson
1–7	103	Shimomura, Iwase, Schuster, Kristjansson

⁴ The Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112-29, 125 Stat. 284, 287–88 (2011), amended 35 U.S.C. § 103, effective March 16, 2013. Because the ’675 patent issued from a patent application filed after this date, the AIA version of § 103 applies. Ex. 1001, codes (21), (22).

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. 7. In its Patent Owner Response, SoundClear initially agrees with Amazon’s contention that no claim terms require construction by stating “no claim construction is necessary.” PO Resp. 4 (citing Ex. 2016 ¶ 43). Despite agreeing with Amazon that no claim construction is necessary, SoundClear identifies two claims terms or phrases that require further construction because, according to SoundClear, “Amazon’s Petition . . . cannot succeed without vitiating the plain and ordinary meanings of [these] two claim phrases, within the context of the claims.” *Id.* at 4–5. We address the two claim terms identified by SoundClear below.

1. “*classify the voice into either a first voice or a second voice*”

In its Patent Owner Response, SoundClear contends that each of independent claims 1, 6, and 7 recite, in relevant part, “calculat[e/ing] a distance between the user and [the/an] output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance.” PO Resp. 5 (citing Ex. 1001, claims 1, 6, 7). SoundClear argues that the ordinary and customary meaning of this claim phrase “requires a classification of *voices*” that “cannot be . . . met by continuous measurements of distance, or distance ranges.” *Id.* According to SoundClear, the specification of the ’675 patent makes clear that the claimed “first voice” and “second voice” “are different classes of *voices*.” *Id.* at 5–6 (citing Ex. 1001, 10:4–8, 10:32–38, 10:47–53). SoundClear further argues that, to the extent the voice classification required by independent claims 1, 6, and 7 is based on a single “feature value,” a person of ordinary skill in the art would have recognized that the feature value threshold must be able to distinguish one class of voice from another. *Id.* at 6 (citing Ex. 2016 ¶¶ 44–46). SoundClear argues that, when user-to-device distance is used as a “feature value,” the “distance threshold must be sufficient for the purpose of classifying the voices.” *Id.* But SoundClear asserts that Amazon’s declarant, Richard Stern, Ph.D., admitted during his deposition that distance may sometimes be sufficient to classify voices, but that is not always the case. *Id.* (citing Ex. 2020, 48:19–49:7).

In its Reply, Amazon provides several reasons as to why it disagrees with SoundClear’s argument that “the received voice cannot be classified based on distance alone” (e.g., “as a ‘near’ or ‘far’ voice depending on whether the distance exceeds a threshold”). Pet. Reply 1 (citing

PO Resp. 5). Amazon argues that SoundClear’s argument in this regard contradicts positions it took in the related district court case between the parties where SoundClear “represented that the ‘first voice’ may be a ‘near/close voice’ and the second voice may be a ‘far/distance voice.’” *Id.* at 1–2 (citing Ex. 1212 ¶¶ 63, 78). Amazon also argues that the plain language of the claims and the specification both support performing classification based on distance alone. *Id.* at 2 (citing Ex. 1001, claim 1, 10:32–38, 10:42–46; Ex. 1202 ¶ 6). In addition, Amazon argues that both declarants confirm that the challenged claims require classifying the voice based on distance alone. *Id.* (citing Ex. 2016 ¶¶ 39, 40, 45; Ex. 1202 ¶ 5).

In its Sur-reply, SoundClear reiterates its argument that the claim language at issue “requires classification of *voices*” that cannot be met by “[a]n arbitrary threshold based on continuous distances measurement, set with no regard to the input voice (the object that must be classified).”

PO Sur-reply 1–2. SoundClear contends that, even when equating “classification” and “measurement,” Amazon’s Reply argument that “*voice classification* can be a distance measurement” fails because “a ‘voice’ cannot be based solely on ‘distance.’” *Id.* at 2 (citing Pet. Reply 1–3).

SoundClear argues that Amazon’s declarant, Dr. Stern, admitted in a related IPR that the claim term “voice” is distinct from the claim term “distance.” *Id.* at 3 (citing Ex. 2020, 28:12–18, 31:11–32, 54:1–6). Moreover, SoundClear asserts that “not *all* distance thresholds are sufficient to distinguish one class of voice from another.” *Id.* (citing Ex. 2020, 48:19–49:7). SoundClear also reiterates that, “[t]o the extent a distance threshold is used as the sole metric to classify voices, such a threshold must be set in a

manner fit for that purpose—the threshold cannot be consistent with more than one class of voices.” *Id.* (citing PO Resp. 6).

Based on the complete record, the parties dispute whether the voice classification required by independent claims 1, 6, and 7 can be based on distance alone. But there is nothing ambiguous or linguistically obscure about the scope and meaning of the aforementioned claim phrase “calculat[e/ing] a distance between the user and [the/an] output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance” that would warrant deviating from its ordinary and customary meaning. *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 phrase] as used in the present claim.”).

In our view, the plain language of this claim phrase controls because it merely requires classifying a voice spoken by a user that is acquired by a device 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, 619–20 (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. 28:25–29:8 (upon inquiring whether the plain language of the challenged claims was controlling, counsel for SoundClear stated “[t]hat’s right.”), 11:11–23 (upon inquiring whether the plain language of the challenged claims was controlling, counsel for Amazon stated “I don’t think you need to [go outside the claims], which is why I’m

starting with the claim language and the specification. I think that resolves the issue.”).

The specification of the ’675 patent further reinforces that the plain language of the claim phrase “classify the voice into either a first voice or a second voice based on the calculated distance” 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 42 analyzes the voice V1 acquired by the voice acquiring unit 30, and classifies the voice V1 to either one of a first voice V1A and a second voice V1B.” Ex. 1001, 9:60–64. The specification makes clear that voice classifying unit 42 performs the functions of analyzing the voice acquired by voice acquiring unit 30 to classify the voice as either a first voice or a second voice “by using any method.” *Id.* at 10:22–24. One such method includes using a proximity sensor to measure a distance between user H and output content-control device 1 illustrated in Figure 1, and then using the distance “as a feature value to perform the classification.” *Id.* at 10:32–38. In the particular case where distance is used as the feature value, “voice classifying unit 42 sets a threshold of the feature value [(i.e., distance)], and classifies the voice V1 to either the first voice V1A or the second voice V1B based on whether the feature value [(i.e., distance)] exceeds the threshold.” *Id.* at 10:42–46.

Indeed, the patentee of the ’675 patent had an opportunity during prosecution to further limit the challenged claims by requiring them to perform the claimed voice classification using distance in conjunction with

other feature values, but declined to do so. *See* Ex. 2014, 156–60. Instead, upon entering a request for continued examination, the patentee amended each of the claims that issued as independent claims 1, 6, and 7 of the ’675 patent to require “calculate[e/ing] a distance between the user and [an/the] output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance.” *Id.* (amending independent claims 1, 6, and 7 of the ’675 patent).

Contrary to SoundClear’s claim construction arguments, we need not require the voice classification required by independent claims 1, 6, and 7 to include something more than a distance measurement. *See* PO Resp. 5; PO Sur-reply 1–2. This would rewrite this claim phrase at issue. *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 something more than a distance measurement either ignores or overlooks the plain language of the claims, the specification, and the prosecution history, each of which, as we explain above, support classifying a voice spoken by a user that is acquired by a device 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 decline to construe separately the claim phrase “calculat[e/ing] a distance between the user and [the/an] output-content control device by a proximity sensor to classify the voice into either a first voice or a second voice based on the calculated distance.” Instead, we construe this claim phrase in accordance with its ordinary and customary meaning.

2. “*whisper*”

In its Patent Owner Response, SoundClear contends that “Amazon’s position as to [dependent] claim 5 relies on a de facto construction of whisper to simply mean a ‘range of speech volumes and rates.’” PO Resp. 7 (quoting Pet. 38) (citing Ex. 2016 ¶ 48). SoundClear, however, argues that Amazon provides no basis on which to conclude that a person of ordinary skill in the art reading the ’675 patent would have understood the claim term “whisper” relates to a “range of speech volumes and rates.” *Id.* (citing Ex. 2017, 86:4–3, 89:15–90:13; Ex. 2016 ¶ 49). Instead, SoundClear argues that “[t]he ’675 patent explains that a ‘whisper’ results from an ‘unvoiced voice spoken without vibrating the vocal chords’” *Id.* (quoting Ex. 1001, 10:4–5) (citing Ex. 1001, 10:5–8, 10:17–22). SoundClear then asserts that “a [person of ordinary skill in the art] would have understood that volume detection alone cannot suffice for determination of a whisper.” *Id.* at 8 (citing Ex. 2016 ¶ 50).

In its Reply, Amazon contends that its arguments set forth in the Petition with respect to dependent claim 5 does not rely on a de facto construction of the claim term “whisper” as meaning a “range of speech volumes and rates.” Pet. Reply 4 (citing PO Resp. 7–8). According to Amazon, dependent claim 5 does not require determining whether the voice is a whisper, but instead “[it] requires that the device be configured to determine that ‘the voice’ is ‘the predetermined voice’ in circumstances where the user whispers to the device.” *Id.* at 16–17 (citing Ex. 1001, claim 5; Ex. 1202 ¶ 19). Amazon asserts that “[n]othing in the claim requires the device to *determine* whether the received voice is a whisper.” *Id.* at 17. Amazon further argues that SoundClear’s reliance on the

specification of the '675 patent is misplaced because it does not address the same situation required by dependent claim 5. *Id.* Amazon asserts that SoundClear's arguments regarding dependent claim 5 "are premised on an improperly narrow reading of the claim." *Id.* (citing Ex. 1202 ¶ 20).

In its Sur-reply, SoundClear disagrees with Amazon's argument that dependent claim 5 does not require determining that the voice is a whisper. PO Sur-reply 4 (citing Pet. Reply 16–18). SoundClear argues that Amazon's "strained interpretation is an affront to the plain language of the claim and directly contradicts the specification." *Id.* According to SoundClear, the plain language of dependent claim 5 "requires that the voice classifying unit determine the voice to be the predetermined voice 'when the voice spoken by the user is a whisper.'" *Id.* (quoting Ex. 1001, claim 5). That is, SoundClear asserts that "the claim expressly requires that the voice classifying unit has determined the voice spoken by a user is a whisper." *Id.* SoundClear further argues that the specification supports this reading of dependent claim 5. *Id.* at 4–5 (citing Ex. 1001, 9:58–10:1).

At the outset, we note that "the specification may reveal a special definition given to a claim term by the patentee In such cases, the [patentee's] lexicography governs." *Phillips*, 415 F.3d at 1316 (citing *CCS Fitness, Inc. v. Brunswick Corp.*, 288 F.3d 1359, 1366 (Fed. Cir. 2002)). Although the patentee is free to define the specific claim terms used to describe his or her invention, "this must be done with reasonable clarity, deliberateness, and precision." *In re Paulsen*, 30 F.3d 1475, 1480 (Fed. Cir. 1994).

Upon reviewing the specification of the '675 patent, we agree with SoundClear that it provides a definition of the claim term "whisper." *See*

PO Resp. 7. That is, in the context of describing how voice classifying unit 42 classifies voice V1 as first voice V1A when voice V1 is a whisper, the specification states that “[a] whisper *is* an unvoiced voice spoken without vibrating the vocal chords, but it is not limited to the unvoiced voice spoken without vibrating the vocal chords as long as being classified to the first voice V1A by, for example, a method described later.” Ex. 1001, 10:4–8 (emphasis added). Later, the specification states the following:

The voice classifying unit 42 classifies the voice V1 either to the first voice V1A or the second voice V1B by using a peak frequency that is equal to or higher than a predetermined intensity in the spectrum as a feature value. For example, the voice classifying unit 42 determines the voice as a whisper to classify to the first voice V1A when the peak frequency is equal to or lower than the threshold, and determines the voice as not a whisper to classify to the second voice V1B when the peak frequency is larger than the threshold.

Id. at 10:13–22. By using the verb “is” following “whisper,” the specification sets forth an explicit definition for the claim term “whisper” as “an unvoiced voice spoken without vibrating the vocal chords.” The specification, however, makes clear that a whisper is not limited to this definition as long as classifying a voice as a whisper includes determining if the peak frequency of the voice is equal to or lower than a threshold.

We do not agree with Amazon’s argument that nothing in dependent claim 5 requires determining whether the voice is a whisper. *See* Pet. Reply 16–17. Dependent claim 5 recites “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” Ex. 1001, 22:5–7. The parties do not dispute that there is no antecedent basis for “the predetermined voice” in either independent claim 1 or dependent claim 5, thereby making it difficult

to ascertain the scope and meaning of the larger claim phrase “determin[ing] the voice to be the predetermined voice when the voice spoken by the user is a whisper.”⁵ The specification, however, provides further context for determining when the voice spoken by the user is a whisper by disclosing the following:

The voice classifying unit 42 classifies the voice V1 into the first voice V1A when the voice V1 spoken by the user is a whisper. The output-content control device 1 detects a whisper and, in that case, generates the first output sentence. It is thereby appropriately determined whether the content is preferable not to be known by the people other than the user H, and is possible to suppress the influence thereof appropriately.

Ex. 1001, 19:23–30. Even with the uncertainty surrounding whether “the predetermined voice” recited in dependent claim 5 refers to either “the first voice” or “the second voice” recited in independent claim 1, this aforementioned description in the specification provides support for SoundClear’s reading of dependent claim 5 as requiring that voice classifying unit 42 within output-content control device 1 to determine whether the voice spoken by a user is a whisper. *See* PO Sur-reply 4–5.

In summary, we find that the specification of the ’675 patent sets forth an explicit definition for the claim term “whisper” as “an unvoiced voice spoken without vibrating the vocal chords,” but it also includes a voice with a peak frequency equal to or lower than a threshold. We also take this

⁵ For purposes of this IPR only, Amazon assumes that “the predetermined voice” recited in dependent claim 5 refers to either “the first voice” or “the second voice” previously recited in independent claim 1. Pet. 37–38. SoundClear does not appear to dispute this reading of the claim term “the predetermined voice.”

opportunity to note that we agree with SoundClear that both the plain language of dependent claim 5 and the specification of the '675 patent support requiring the voice classifying to determine whether the voice spoken by a user is a whisper.

B. Obviousness over Shin and Aoyama

Amazon contends that claims 1–7 of the '675 patent are unpatentable under 35 U.S.C. § 103 as obvious over the combined teachings of Shin and Aoyama. Pet. 7–41. Amazon contends that the combined teachings of Shin and Aoyama account for the subject matter of each challenged claim, and 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.

SoundClear contends that the proposed combination of Shin and Aoyama does not teach or suggest limitations 1[b][ii], 1[f], and 1[g]. PO Resp. 23–33, 37; PO Sur-reply 9–19. SoundClear also argues that Shin does not disclose all the components as arranged in independent claim 1. PO Resp. 38. To support its positions, SoundClear submits the Declaration of Joseph C. McAlexander III, P.E., with the Patent Owner Response. *See* Ex. 2016.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1–4, 6, and 7 of the '675 patent are unpatentable as obvious over the combined teachings of Shin and Aoyama. We, however, determine that Amazon has not shown by a preponderance of evidence that claim 5 of the '675 patent is unpatentable as obvious over the combined teachings of Shin and Aoyama. We begin our

analysis with the principles of law that generally apply to an asserted ground based on obviousness, we then assess the level of skill in the art, next we provide overviews of Shin and Aoyama, and then we turn to the parties' contentions with respect to the challenged claims.

1. Principles of Law

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 ¶¶ 28–32). 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 ’675 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.

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 ’675 patent and the asserted prior art, and (2) it is supported by the testimony of Dr. Stern. We note, however, that our 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).

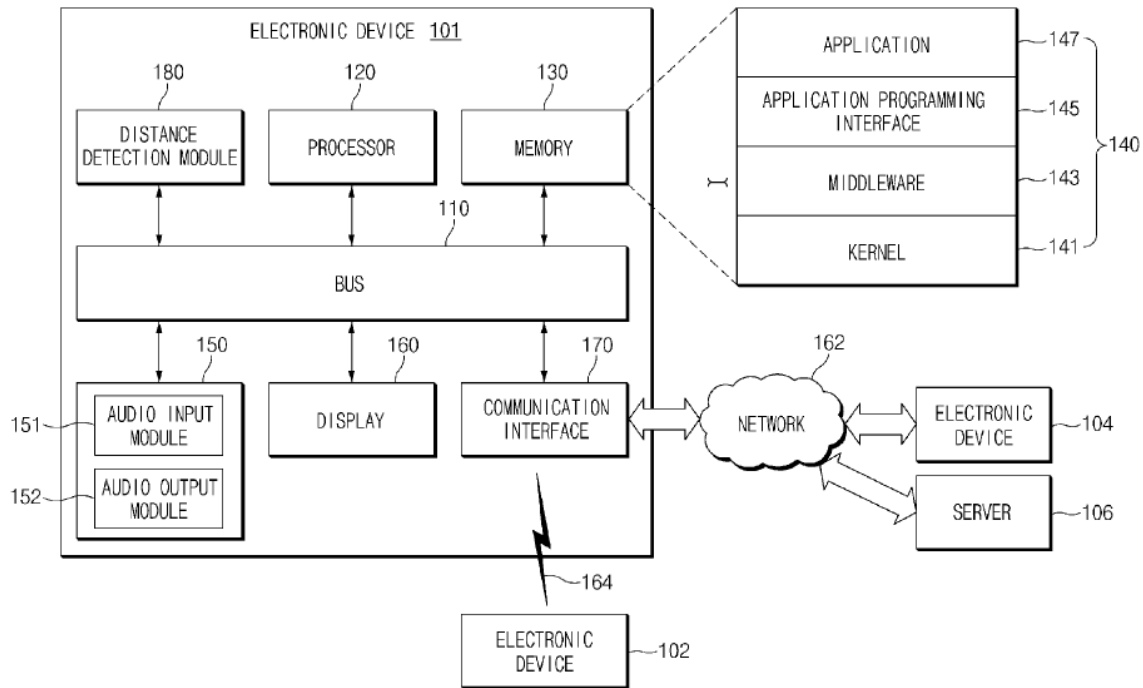


FIG.3

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. *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 or simplified content for a user more than 1 meter away from electronic device 101. *Id.* ¶¶ 86–88, Table 3.

4. Overview of Aoyama

Aoyama generally relates to “an information processing device, an information processing method, and a program.” Ex. 1010 ¶ 1. Aoyama discloses information processing device 1 with controller 10 that further includes, among other things, response parameter generation unit 16 and response generation unit 17. *Id.* ¶¶ 56, 61, Fig. 2. Response generation unit 16 provides “information indicating the name of the response component based on the speech style of the user” to response generation unit 17, which, in turn, “generate[s] a response to the sound input [received] from the user.” *Id.* ¶¶ 131, 132. As one example, in response to a user stating “[c]heck my schedule tomorrow at 1 pm,” response generation unit 17 generates the response “[y]ou have a meeting with Mr. Yamada in room A.” *Id.* ¶ 138. Notably, in this example, response generation unit 17 replaces “the personal name ‘Taro Yamada’ with the name ‘Mr. Yamada’ in accordance with the speech style of the user.” *Id.*; *see also id.* ¶¶ 139, 140 (disclosing other examples where response generation unit 17 replaces a specific date (i.e., changing “10/1/2014” to “today,” “tomorrow,” or “the day after tomorrow”) or it replaces a specific time (i.e., changing “13:00” to “1 pm”)).

5. Claims 1, 6, and 7

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. 7–29, 76. For the most part, the same arguments and evidence Amazon provides to explain how the combined teachings of Shin and Aoyama account for all the limitations of independent claim 1 equally apply

to the corresponding limitations of independent claims 6 and 7. *See id.* at 40–41.

The preamble of independent claim 1 recites 1[pre]—“[a]n output-content control device.” Ex. 1001, 20:60. To the extent the preamble should be treated as limiting, Amazon contends that Shin teaches or suggests the features of the preamble because its electronic device (e.g., electronic device 101 illustrated in Figure 1) receives voice input from a user, “generate[s] content corresponding to a result of analyzing the voice input,” and “provide[s] the generated content as sound.” Pet. 7 (quoting Ex. 1003 ¶ 37) (citing Ex. 1003, Abstract, ¶¶ 3, 8–11, 38, 39, 48–60, 69–72, claims 1, 15, 20) (alterations in original). Amazon argues that Shin’s electronic device 101 may be a smartphone, PC, or home appliance. *Id.* at 8 (citing Ex. 1003 ¶ 31).

Independent claim 1 recites limitation 1[a]—“a voice acquiring unit configured to acquire a voice spoken by a user.” Ex. 1001, 20:61–62. Amazon contends that Shin teaches or suggests this limitation because electronic device 101 includes, among other things, processor 120 and audio input module 151 “implemented with a microphone” that work together to “obtain a user’s speech as voice input.” Pet. 8–9 (quoting Ex. 1003 ¶ 62) (citing Ex. 1003 ¶¶ 3–11, 29, 37–39, 41, 45, 48, 50–53, 60, 61, 155, 158, Fig. 3; Ex. 1002 ¶¶ 50–53).

Independent claim 1 recites limitations 1[b]—“a voice classifying unit configured to 1[b][i] calculate a distance between the user and the output-content control device by a proximity sensor to 1[b][ii] classify the voice into either a first voice or a second voice based on the calculated distance.” Ex. 1001, 20:63–67. Amazon contends that Shin teaches or suggests

limitations 1[b] and 1[b][i] because processor 120 and distance detection module 180, which may include “a proximity sensor,” work together to calculate a distance between the user and electronic device 101. Pet. 10–11 (quoting Ex. 1003 ¶ 153) (citing Ex. 1003 ¶¶ 42, 45, 53, 54, 66, 75, 77, 96, 143, claims 5, 9, Figs. 3, 5A; Ex. 1002 ¶¶ 56–58). Amazon contends that Shin teaches or suggests limitation 1[b][ii] because processor 120 executes a speech recognition application that processes the voice input received from the user and generates output based on the distance between the user and electronic device 101. *Id.* at 12 (citing Ex. 1003 ¶¶ 37, 39, 47, 53, 77, 78, 86; Ex. 1002 ¶ 60). Amazon argues that an example of Shin’s processor 120 classifying the voice 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
Second Voice	$D < 1 \text{ m}$	Amount of information of original content
First Voice	$1 \text{ m} \leq D < 2 \text{ m}$	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. at 13 (citing Ex. 1003, Table 3). Amazon argues that the consolidated, annotated version of Shin’s Table 3 illustrates classifying the user’s voice input as a “first voice” when the distance between the user and electronic device 101 is 1 to 2 meters, and classifying the user’s voice input as a “second voice” when the distance between the user and electronic device 101 is less than 1 meter. *Id.* at 13–14 (citing Ex. 1003 ¶¶ 51, 86–87, Table 3; Ex. 1002 ¶¶ 59–62).

Independent claim 1 recites limitation 1[c]—“an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by the user.” Ex. 1001, 21:1–4. Independent claim 1 also recites limitation 1[d]—“a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information.” *Id.* at 21:5–8. Amazon contends that Shin teaches or suggests limitations 1[c] and 1[d] because processor 120 executes a speech recognition application that processes the user’s voice input and generates corresponding output by analyzing the user’s request. Pet. 16 (citing Ex. 1003 ¶¶ 37, 48, 58–60, 62, 67, 71, 84, 88, 89, 103, 107, 120, 126, 127, 142; Ex. 1002 ¶ 73). Amazon further argues that Shin teaches analyzing the acquired voice “to detect intention information” and to acquire “notification information . . . based on the intention information” because it discloses examples of analyzing the user’s voice input and then acquiring notification based on the user’s intent. *See id.* at 16–17. As one example, Shin discloses analyzing the user’s voice input of “Let me know today’s weather” to determine the weather and to acquire information such as sky conditions, air temperature, or rainfall. *Id.* at 16 (quoting Ex. 1003 ¶ 88). As another example, Shin discloses analyzing a user’s voice input of “Let me know what time it is now” to determine the time and to acquire the current time. *Id.* at 16–17 (quoting Ex. 1003 ¶ 37) (citing Ex. 1003 ¶ 39; Ex. 1002 ¶ 74). Amazon argues that Shin’s electronic device 101 either stores an application in memory that “analyz[es] a voice input[] and generat[es] and/or provid[es] the content corresponding to the voice input,” or delegates some processing

to another device to deliver results. *Id.* at 17 (quoting Ex. 1003 ¶ 60) (citing Ex. 1003 ¶¶ 64, 65, 67, 174, 175; Ex. 1002 ¶ 75) (alterations in original).

Alternatively, Amazon contends that, even if Shin does not teach or suggest limitations 1[c] and 1[d], Aoyama teaches or suggests these two limitations. Pet. 18–19. Amazon argues that Aoyama teaches or suggests limitation 1[c] because it discloses a meaning analysis unit 114 that uses natural language processing to determine the “meaning” of a user’s voice input and to specify a “task” to be executed. *Id.* at 18 (first quoting Ex. 1010 ¶ 72, then quoting Ex. 1010 ¶ 74) (citing Ex. 1010 ¶¶ 73, 75–84, 171–173, 230–232, Figs. 3, 8, 13). As one example, Amazon argues that Aoyama’s meaning analysis unit 114 receives a user’s voice input of “Check my schedule for tomorrow at 1 pm” and then detects the intended scheduled information. *Id.* (quoting Ex. 1010 ¶ 76) (citing Ex. 1010 ¶¶ 77–79, 170–173). Amazon also argues that Aoyama teaches or suggests limitation 1[d] because it discloses response component generation unit 12 that “executes the task indicated by the information acquired from the meaning analysis unit 114 . . . and acquires information necessary to generate a response to the user.” *Id.* at 19 (quoting Ex. 1010 ¶ 127) (citing Ex. 1010 ¶¶ 128–130, 174–176, 233, Figs. 3, 8, 13). For instance, in the example where the user’s voice input requests scheduling information, Aoyama discloses that response component generation unit 12 “acquires a search result” from a scheduling application. *Id.* (quoting Ex. 1010 ¶ 129).

Turning to rationale to combine and reasonable expectation of success, Amazon contends a person of ordinary skill in the art would have been motivated to include Aoyama’s meaning analysis unit 114 and response component generation unit 12 in Shin’s electronic device 101 for

several reasons. Pet. 19 (citing Ex. 1002 ¶¶ 82–84). As one example, Amazon argues that the proposed combination would have reflected the addition of one known element (i.e., “Aoyama’s natural language processing and information retrieval”) to another known element (i.e., “Shin’s voice processing and response device”) to obtain predictable results (i.e., “identifying a user’s intent and providing information responsive to such intent”). *Id.* at 19–20 (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶ 83). According to Amazon, there would have been a reasonable expectation of success in modifying Shin in this way “because doing so would have been trivial, involving at most minor changes to Shin’s system to implement Aoyama’s units.” *Id.* at 20 (citing Ex. 1002 ¶ 84).

Independent claim 1 recites limitation 1[e]—“an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information.” Ex. 1001, 21:9–11. Relying on its arguments and evidence presented for limitations 1[b], 1[b][i], 1[b][ii], 1[c], and 1[d], Amazon contends that that Shin teaches or suggests limitation 1[e] because it discloses that processor 120 generates content for output to the user. Pet. 21. For example, Amazon argues that, in the example where the user’s voice input requests the time of day, Shin’s processor 120 generates the output sentence “The current time is nine ten AM.” *Id.* (quoting Ex. 1003 ¶ 37). Amazon further argues that Shin’s text-to-speech technology converts the output sentence to speech. *Id.* (citing Ex. 1003 ¶¶ 7, 37, 39, 48, 62, 72, 109, 131, 139, 158; Ex. 1002 ¶ 87). Relying on its arguments and evidence presented for limitation 1[d], Amazon also contends that Shin’s processor 120 analyses the user’s voice input to acquire notification information according to the user’s interest. *Id.*

Amazon argues that Shin’s notification information is incorporated into an output sentence, such as providing sky conditions, air temperature, or rainfall in the context of the weather example. *Id.* (citing Ex. 1003 ¶¶ 37, 88; Ex. 1002 ¶ 88).

Independent claim 1 recites limitation 1[f]—“wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word.” Ex. 1001, 21:9–17. Amazon contends that the proposed combination of Shin and Aoyama teach or suggest limitation 1[f] because Shin’s processor 120 adjusts the amount of output information provided to the user based on the distance between the user and electronic device 101. Pet. 22 (citing Ex. 1003 ¶¶ 51, 53, 86–90, claims 8, 17). Amazon argues that Shin discloses, when the user’s “distance increases by 1 m, the electronic device 101 may decrease the output amount of information by making the abbreviation level of the original content high.” *Id.* (quoting Ex. 1003 ¶ 87) (citing Ex. 1003 ¶ 86). Amazon further argues that Shin discloses, when the user’s distance is 1 to 2 meters, electronic device 101 generates “[c]ontent of first abbreviation level,” which may be described more “simply” and abbreviated to include “a gist of the response,” with less information. *Id.* at 23 (first quoting Ex. 1003 ¶ 86, then quoting Ex. 1003 ¶ 51) (citing Ex. 1002 ¶ 92).

Amazon then turns to Aoyama to more clearly teach or suggest generating Shin’s abbreviated or simplified content by replacing words. Pet. 23 (citing Ex. 1002 ¶¶ 93–103). Amazon argues that Aoyama’s device (e.g., a smartphone) recognizes and analyzes a user’s speech, executes

processing based on the voice input, and audibly presents the result as a sentence. *Id.* (citing Ex. 1010 ¶¶ 42, 51, 127, 161, 241–243, 253, Fig. 8). Amazon asserts that Aoyama’s device tailors output by replacing words based on the user or situation, such as by replacing a user’s formal name with an informal name. *Id.* at 23–24 (citing Ex. 1010 ¶¶ 120–124, 138–140, 177, 218, 234–236, Figs. 7, 14; Ex. 1002 ¶ 94).

Turning to rationale to combine and reasonable expectation of success, Amazon contends a person of ordinary skill in the art would have been motivated to configure Shin’s processor 120 to adjust output content based on the user’s distance from electronic device 101 by incorporating Aoyama’s word replacement feature for several reasons. Pet. 25 (citing Ex. 1002 ¶¶ 96–103). As one example, Amazon argues that Aoyama’s word replacement feature furthers Shin’s goal to abbreviate or simplify content output to users at a greater distance from electronic device 101 to make the content easier for the users to grasp. *Id.* According to Amazon, this may include replacing full names with simpler nicknames and replacing complicated date and time formats with simpler terms, each of which would make the content output to the user easier to understand at greater distances. *Id.* (citing Ex. 1010 ¶¶ 138–140, 221; Ex. 1002 ¶ 97). As another example, Amazon argues that the proposed combination merely represents the addition of one known element (i.e., “Aoyama’s word replacement” feature) to another known element (i.e., “Shin’s distance-based content abbreviation/simplification”) to obtain predictable results (i.e., “using word replacement to simplify content for distant users”). *Id.* at 25–26 (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶ 100). According to Amazon, there would have been a reasonable expectation of success in making this proposed

combination because (1) configuring Shin’s processor 120 accordingly “would have been trivial, involving straightforward changes to enhance the processor’s approach to simplifying and abbreviating its output”; and (2) this approach was well-known as evidenced by Ha’s disclosure of replacing “team meeting” and “date” with “schedules” based on the user’s environment.⁶ *Id.* at 27–28 (first quoting Ex. 1011 ¶ 97, then quoting Ex. 1011 ¶ 98) (citing Ex. 1002 ¶ 102; Ex. 1011, Fig. 7).

Independent claim 1 recites limitation 1[g]—“wherein the output-content generating unit is further configured . . . to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information.” Ex. 1001, 21:11–20. Amazon contends that Shin teaches or suggests limitation 1[g] because it discloses that, when classifying a voice as a second voice because the user is less than 1 meter away from electronic device 101, electronic device 101 generates a second output comprising all available information (i.e., by outputting unabbreviated content). Pet. 28–29 (citing Ex. 1003 ¶¶ 51, 87, 88, Table 3; Ex. 1002 ¶ 106). Amazon argues that, in the example of outputting the time as “The current time is nine ten AM,” the word “time” conveys all the intention information derived from the user statement “Let me know what time it is now.” *Id.* at 29 (quoting Ex. 1003 ¶ 37) (citing Ex. 1003 ¶¶ 39, 107). Amazon asserts that the words “nine ten” represent all content information determined for that user statement. *Id.*

⁶ We understand Amazon to introduce Ha as evidence of the general background knowledge in the art pertaining to the use of a word replacement by output-content control devices.

Notwithstanding SoundClear’s arguments to the contrary, Amazon has shown by a preponderance of the evidence that the combined teachings of Shin and Aoyama account for all the limitations of independent claims 1, 6, and 7. During trial, SoundClear contends that the proposed combination of Shin and Aoyama does not teach or suggest limitations 1[b][ii], 1[f], and 1[g]. PO Resp. 23–33, 37; PO Sur-reply 9–19. SoundClear also argues that Shin does not disclose all the components as arranged in independent claim 1. PO Resp. 38. We begin our analysis by addressing SoundClear’s arguments directed to limitations 1[b][ii], 1[f], and 1[g], and then we address SoundClear’s arguments directed to whether Shin discloses the claimed components as arranged in independent claim 1.

a. Limitation 1[b][ii]—“a voice classifying unit configured to . . . classify the voice into either a first voice or a second voice based on the calculated distance”

In its Patent Owner Response, SoundClear contends that Amazon only relies on Shin to teach or suggest limitation 1[b][ii]. PO Resp. 23 (citing Ex. 1001, claims 1, 6, 7; Pet. 12–15). SoundClear, however, argues that “Shin discloses determining output schemes based on distance ranges, it does not disclose that any such range is sufficient for the purpose of classifying voices.” *Id.* at 23–24 (citing Ex. 2017, 65:10–22; Pet. 22). SoundClear asserts that Amazon does not identify an explicit disclosure in Shin of “basing *any* function or output on ‘voice.’” *Id.* at 24 (citing Pet. 12). 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 voice classification. *Id.* at 24 (citing Ex. 2016 ¶¶ 82, 83). SoundClear also argues that Shin only mentions “classification” of “content”

in the context of output—not voice input. *Id.* (citing Ex. 1003 ¶ 51).

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.” PO Resp. 25–26 (citing Ex. 2016 ¶¶ 84–87; Pet. 13; Ex. 1003, Fig. 3).

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 disclosure of distance between the user and the device to be synonymous with voice or classification. *Id.* at 26 (citing Ex. 1003 ¶¶ 53, 87; Ex. 2016 ¶ 88). Stated differently, SoundClear argues that Shin merely correlates content output to distance between the user and the device, but it does not classify the voice input in any way, let alone classify the voice input as a first or second voice. *Id.* 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 device. *Id.* at 27–28 (citing Ex. 2020, 23:5–24:17, 72:12–73:8; Ex. 2017, 49:7–15; Ex. 2016 ¶¶ 90, 91).

In its Reply, Amazon contends that Shin is replete with disclosures of analyzing voice input to determine an output scheme. Pet. Reply 10 (citing Ex. 1003, Abstract, ¶¶ 48, 49, 53, Fig. 5A; Ex. 1202 ¶ 11). Amazon further argues that SoundClear’s argument about Shin lacking an explicit disclosure of classification is without 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 11. 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 11 (citing Ex. 1003, Table 3; Pet. 12–14; Ex. 1202 ¶ 12). 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 device. *Id.* at 12. According to Amazon, Dr. Stern testified that Shin’s Table 3 discloses voice classification. *Id.* at 12–13 (citing Ex. 2020, 24:9–17, 68:16–69:18, 70:16–71:6, 71:18–72:7).

In its Sur-reply, SoundClear contends that Amazon does not disclose classifying a “voice” because distance thresholds alone “do not . . . delineate one class of voice from another.” PO Sur-reply 9. According to SoundClear, Amazon has not demonstrated that Shin’s distance thresholds are capable of classifying a voice, nor has Amazon shown that a person of ordinary skill in the art would have reasonably understood Shin to teach or suggest actually doing so. *Id.* Bottom line, SoundClear asserts that “[t]he Petition is devoid of any reasoned analysis” that Shin “*actually* classifies the voice.” *Id.* at 9–10. According to SoundClear, Dr. Stern provides deposition testimony confirming that Shin “merely discloses measuring variables (e.g., distance) and modifying output content based on the measurement,” but never discloses voice classification. *Id.* at 10–13 (citing PO Resp. 12–13; Ex. 2017, 57:6–13, 57:18–58:5, 58:6–12, 60:13–61:4, 95:7–17, 119:13–19; Ex. 2021, 21:8–22:16, 33:5–34:15, 62:22–63:8, 64:5–65:2, 65:6–12, 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 Shin discloses voice classification. *See id.* at 13. At best, SoundClear asserts that Shin discloses classifying distances rather than classifying voices, as required by the challenged claims. *Id.* at 14–15 (citing Pet. Reply 11; Ex. 2017, 48:3–7, 49:7–15, 117:1–21, 118:2–13, 119:6–22).

Based on the complete record, we agree with Amazon that Shin teaches or suggests 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][ii]. Pet. 12–15; Pet. Reply 10–13. 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 high.” *Id.*

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 1 to 2 meters, and another that classifies the user’s voice as a second voice when the distance between the user and electronic device 101 is less than 1 meter. *See* Pet. 13 (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][ii].

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). When conducting an obviousness analysis, we must consider Shin “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, Shin’s processor 120 performs the voice classification required by limitation 1[b][ii] 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. 25–28; PO Sur-reply 10–13. 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][ii]. *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 output amount of 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.

b. Limitation 1[f]—“the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word”

In its Patent Owner Response, SoundClear contends that the proposed combination of Shin and Aoyama does not teach or suggest limitation 1[f]. PO Resp. 28–29 (citing Ex. 1001, claim 1, 6, 7). SoundClear argues that Shin's disclosure of abbreviating or simplifying content does not include

“replacing words.” *Id.* at 29 (citing Pet. 22–23; Ex. 2016 ¶¶ 94, 95).

That is, SoundClear asserts that “Shin discloses abbreviating content by excluding certain information included in the original content,” but “Shin nowhere disclose abbreviating content by replacing words.” *Id.* at 29–30 (citing Ex. 1003 ¶ 88; Ex. 2016 ¶¶ 96, 97). SoundClear further argues that Aoyama does not render replacing words obvious because Aoyama merely “teaches inputting words into a response based on the ‘speech style of the user,’ but “[i]t does not make any adjustment based on whether the user’s voice is determined to be a ‘first voice’ or ‘second voice,’ as claimed.” *Id.* at 30 (quoting Ex. 1010 ¶ 121) (citing Pet. 23–25; Ex. 2016 ¶ 98). As one example, SoundClear argues that Aoyama may refer to the name “Taro Yamada” as “Mr. Yamada,” “Taro,” or “Taro Yamada” based on the user’s unique expression for personal names, but this does not disclose the claimed voice classification, nor does it provide any basis for word replacement based on voice classification. *Id.* at 30–31 (citing Ex. 1010 ¶¶ 86–89, 121–124; Ex. 2017, 21:7–13; Ex. 2016 ¶ 98).

SoundClear also contends that Aoyama’s word choices based on the “speech style of the user” and its change in “amount of information” (“similar to Shin’s content abbreviation”) is based on the user’s “degree of composure,” but these two features “involve separate and distinct processes and metrics.” PO Resp. 31 (first quoting Ex. 1010 ¶ 121, then quoting Ex. 1010 ¶ 145) (citing Ex. 1010 ¶¶ 122–124, 146, 147; Ex. 2016 ¶ 99); *see also id.* at 32–33 (arguing the differences between Aoyama’s disclosure of inputting words based on the “speech style of the user” versus changing the “amount of information output as a response” based on a user’s “degree of composure”). As a result, SoundClear argues that a person of ordinary skill

in the art would not have modified Shin to include Aoyama's word replacement feature in Shin's output content based on user-to-device distance. *Id.* at 31–32. Instead, SoundClear asserts that a person of ordinary skill in the art would have understood that Aoyama teaches away from replacing words based on user-to-device distance because “Aoyama teaches that words are only to be chosen in response to ‘speaking style’” and “[i]t would led to a response with which ‘the user may feel uncomfortable.’” *Id.* at 32 (quoting Ex. 1010 ¶ 49, then quoting Ex. 1010 ¶ 49) (citing Ex. 2016 ¶ 99).

In its Reply, Amazon maintains that the proposed combination of Shin and Aoyama teaches or suggests replacing words. Pet. Reply 13. According to Amazon, contrary to SoundClear's argument, a person of ordinary skill in the art would have understood that Shin's disclosure of abbreviating or simplifying content includes replacing words in the output. *Id.* at 13–14 (citing Pet. 22–23; Ex. 1002 ¶ 92; Ex. 1207, 2; Ex. 1202 ¶ 14). Amazon also argues that, although Shin discloses an example of abbreviating content by excluding certain portions of the output content, Shin is not limited “to *only* abbreviating output by excluding words rather than replacing words.” *Id.* at 14 (citing Ex. 1202 ¶ 15). Amazon also contends that SoundClear's attempt to dismiss Aoyama's word replacement feature fails for the following two reasons: (1) Amazon does not rely on Aoyama for voice classification, but rather Amazon relies on Shin to teach or suggest this claimed feature; and (2) Aoyama's system provides at least two options for replacing words, but they are irrelevant here because Amazon relies on Shin's voice classification to determine when to replace words. *Id.* (citing PO Resp. 30–32; Pet. 10–15, 22–25). Lastly, Amazon asserts that SoundClear's remaining argument

regarding motivation to combine the teachings of Shin and Aoyama are wrong for the following three additional reasons: (1) “Aoyama does not teach that word replacement occurs *only* in response to speaking style,” but “rather . . . Aoyama teaches speaking style as one option for determining word replacement”; (2) merely because Aoyama teaches accomplishing something one way (e.g., word replacement based on speaking style) would not discourage a person of ordinary skill in the art from pursuing other ways (e.g., word replacement based on distance); and (3) Aoyama must be considered for everything it teaches—not just for an embodiment that is similar to Shin’s system that adjusts the amount of information output to the user. Pet. Reply 15–16 (citing Ex. 1010 ¶¶ 128, 162, 255, 256; Ex. 1202 ¶¶ 17, 18; Pet. 25–28).

In its Sur-reply, SoundClear maintains that neither Shin nor Aoyama discloses replacing words as required by limitation 1[f]. PO Sur-reply 16. Beginning with Shin, SoundClear contends that Amazon’s reliance on a dictionary definition for the word “abbreviation” merely indicates that Shin “discloses abbreviation of content” by excluding certain words—not “abbreviation of sets of words” by replacing words. *Id.* (citing Ex. 1207, 2). SoundClear asserts that Amazon does not—and cannot—cite to a disclosure in Shin that “*actually* discloses word replacement.” *Id.* at 17. SoundClear argues that Dr. Stern’s reply testimony fares no better because he does not cite anything to support that Shin teaches or suggests replacing words in its content output to the user. *Id.* (citing Ex. 1202 ¶ 15). Turning to Aoyama, SoundClear maintains its argument that “Aoyama does not disclose word replacement based on voice classification.” *Id.* (citing PO Resp. 30–31). Lastly, SoundClear maintains its argument that a person of ordinary skill in

the art would not have been motivated to modify Shin to include Aoyama's word replacement feature for the following reasons: (1) Aoyama teaches speaking style as the only option for replacing words; (2) Aoyama teaches away from replacing words based on distance because doing so would result in that which Aoyama seeks to avoid—namely, generating a response with which “the user may feel uncomfortable”; and (3) Aoyama contemplates modifying the output and replacing words, but it “does so for wholly different purposes under different circumstances.” *Id.* at 17–18 (quoting Ex. 1010 ¶ 49) (citing Pet. Reply 14–15; PO Resp. 32–33).

Based on the complete record, we agree with Amazon that the proposed combination of Shin and Aoyama teaches or suggests, after classifying the voice as a first voice, generating a first output sentence that replaces one word with another word, as required by limitation 1[f]. Pet. 22–28; Pet. Reply 13–16. As we explain above, 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.5.a. More specifically, 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 1 to 2 meters. *See id.*

Shin further discloses that the output scheme associated with this classification level includes the following: 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.” Ex. 1003 ¶ 87. This

means that, when classifying the user's voice as a first voice because the distance between the user and electronic device 101 is 1 to 2 meters, Shin discloses that processor 120 determines an output scheme that outputs "content of first abbreviation level." *Id.* ¶ 86. Shin discloses that abbreviated or abstract content "includes only the gist of the response corresponding to the voice input" where "the content is described most simply." *Id.* ¶ 51.

Amazon's declarant, Dr. Stern, testifies that "[a] person of ordinary skill in the art would have readily understood Shin's disclosure of abbreviating content to describe it more simply or provide the 'gist' includes replacing words." Ex. 1002 ¶ 92. Later, when explaining why it would have been obvious to a person of ordinary skill in the art to configure Shin's processor 120 to include Aoyama's word replacement feature, Dr. Stern testifies "Shin's device presents only two design options for abbreviating or simplifying its output [to the user]: omitting words or replacing them." *Id.* ¶ 101. We credit Dr. Stern's declaration testimony that Shin's abbreviation or simplification of content output to the user includes omitting or replacing words because it is consistent with the disclosures in Shin that we identify above. *See, e.g.*, Ex. 1003 ¶¶ 51 (disclosing that "processor 120 may vary or adjust an output amount of information by extracting or reconfiguring the content"), 86–88 (disclosing that processor 120 "determines an output scheme . . . such that an output amount of information of the content decreases as a distance between the user and the electronic device 101 increases," including, as one example, by omitting words), Fig. 3.

In any event, even if we were to treat Shin's abbreviation or simplification of content output to the user as limited to omitting words

instead of replacing words, Aoyama's word replacement feature would fill this gap. Similar to Shin's electronic device 101, Aoyama's information processing device 1 recognizes content spoken by a user received as sound input, analyzes the content using sound recognition technology or natural language processing technology, and then executes various processes with the content to present a natural sentence as sound or text to the user.

Ex. 1010 ¶¶ 42. Aoyama disclose that information processing device 1 includes, among other things, response parameter generation unit 16 that provides "information indicating the name of the response component based on the speech style of the user" to response generation unit 17, which, in turn, "generate[s] a response to the sound input [received] from the user." *Id.* ¶¶ 131, 132. Notably, Aoyama's response generation unit 17 tailors a response based on the voice input received from the user. *Id.* ¶¶ 120–124, Fig. 7; *see also id.* ¶ 218 (disclosing that "information processing device 1 may reproduce the switch between names (that is, a speech style) in accordance with [different] situations").

Figure 7 of Aoyama, reproduced below, illustrates a diagram of an example of providing a response parameter based on "the speech style of the user." Ex. 1010 ¶¶ 19, 120.

FIG.7

User ID : 000001

PERSONAL NAME DATA

FORMAL NAME	NAME
MAIKO TANAKA	MAIKO
YUKARI SUZUKI	YUKARI-CHAN
TARO YAMADA	MR. YAMADA

DATE-TIME EXPRESSION PATTERN

WESTERN CALENDAR / JAPANESE CALENDAR	JAPANESE CALENDAR MODEL
12:00/24:00	12-HOUR MODEL WITH AM AND PM

⋮

Figure 7 illustrates response parameter d10 indicating the “speech style of the user” for “User ID: 000001” that includes both personal name data d11 and date-time expression pattern d13. *Id.* ¶ 121. As one example of using the personal name data d11 illustrated in Figure 7, Aoyama discloses that response parameter generation unit 16 receives notification of the name corresponding to “Taro Yamada” from response generation unit 17, response parameter generation unit 16 extracts the response parameter d10 that applies to “Taro Yamada,” response parameter generation unit 16 notifies response generation unit 17 that a user with the name “Taro Yamada” should be called “Mr. Yamada” in accordance with the user’s speech style, and then response generation unit 17 generates a response of “Mr. Yamada.” *Id.* ¶ 124. Aoyama discloses how response generation unit 17 uses response parameters like the ones illustrated in Figure 7 in several different situations, including an inquiry about a user’s schedule, a request for date and time, etc. *Id.* ¶¶ 138–140.

Based on these cited disclosures, Amazon argues—and we agree—that Aoyama teaches or suggests that information processing device 1 tailors the content output to the user by replacing words. *See* Pet. 23–25. As one example pertaining to a user’s name, we find that Aoyama’s information processing device 1 is capable of receiving the name “Taro Yamada” and then replacing “Taro” with “Mr.” so the resulting name output to the user is “Mr. Yamada.”

Taking this teaching or suggestion from Aoyama and applying it to Shin, Amazon argues—and we agree—that a person of ordinary skill in the art would have configured Shin’s processor 120 that classifies the user’s voice based on distance thresholds between the user and electronic device 101 by incorporating Aoyama’s word replacement feature to generate an output sentence that replaces one word with another word, as required by limitation 1[f]. More specifically, after classifying the user’s voice as a first voice because the distance between the user and electronic device 101 is 1 to 2 meters, we find that Shin’s processor 120 is capable of outputting content to the user at a first abbreviation level by using Aoyama’s word replacement feature, thereby generating an output sentence that replaces one word with another word (e.g., by generating an output sentence that takes the name “Taro Yamada” and replaces “Taro” with “Mr.” so the resulting name output to the user is “Mr. Yamada”). In our view, configuring Shin’s processor 120 in this way compliments Shin’s goal of abbreviating or simplifying the content output to a user at a greater distance from electronic device 101 to make the content easier for the user to understand. *See, e.g.*, Ex. 1003 ¶ 89 (disclosing that “a user who requests content with a voice input to the electronic device 101 while being far away from the electronic device 101

may want core content corresponding to the voice input rather than detailed content”).

We do not agree with SoundClear’s argument that “speech style of the user” is the only option Aoyama provides for replacing words. *See* PO Resp. 30; PO Sur-reply 17–18. Although SoundClear correctly notes that Aoyama’s response generation unit 17 generates a response based on “the speech style of user,” Aoyama indicates this is not the only option for replacing words. Ex. 1010 ¶ 121. As one example, Aoyama also discloses that “response generation unit 17 may control an amount of information output as the response in accordance with the acquired information indicating the ‘degree of composure.’” *Id.* ¶ 145. Aoyama discloses that “degree of composure” includes generating different responses depending on whether “the user is in a hurried state” or whether “the user is calm.” *Id.* ¶ 147. As another example, Aoyama discloses that information processing device 1 may vary its process for generating a response depending on the situation at hand. *Id.* ¶ 218; *see also id.* ¶¶ 138–140 (disclosing how response generation unit 17 uses response parameters like the ones illustrated in Figure 7 in several different situations, including an inquiry about a user’s schedule, a request for date and time, etc.).

Nor do we agree with SoundClear’s argument that Aoyama does not disclose the claimed voice classification, much less provide any basis for word replacement based on voice classification. *See* PO Resp. 30–31; PO Sur-reply 17. SoundClear’s argument improperly attacks Aoyama in isolation. It is well settled that “non-obviousness [cannot be established] by attacking references individually,” when, as here, the asserted obviousness ground is based upon the combined teachings of Shin and Aoyama. *In re*

Keller, 642 F.2d 413, 426 (CCPA 1981). Instead, the test is what the combined teachings of these references would fairly suggest to a person of ordinary skill in the art. *In re Young*, 927 F.2d 588, 591 (Fed. Cir. 1991). As we explain above, Amazon’s proposed obviousness ground relies on Shin—not Aoyama—to teach or suggest the voice classification required by limitation 1[b][i]. *See* Pet. 12–15. Amazon turns to Shin’s voice classification together with Aoyama’s word replacement feature to teach or suggest limitation 1[f]. *See id.* at 22–28.

Lastly, we do not agree with SoundClear’s argument that a person of ordinary skill in the art would not have been motivated to combine the teachings of Shin with those of Aoyama to arrive at limitation 1[f]. *See* PO Resp. 31–33; PO Sur-reply 17–18. To support its assertion that Aoyama teaches away from replacing words based on distance, SoundClear directs us to Aoyama’s statement that “the user may feel uncomfortable.” Ex. 1010 ¶ 49. This is misplaced because SoundClear fails to recognize that this cited disclosure in Aoyama refers to a user receiving a response from information processing device 1 in a speaking style different from the user’s preferred speaking style (e.g., receiving a response of “15:00” rather than a response of “3 pm”). *Id.* As we explain above, Aoyama provides a solution to this problem by providing several different options for replacing words, including by providing a response based on “the speech style of user,” based on the “degree of composure” (e.g., calm, hurried, etc.), or depending on the situation at hand. *Id.* ¶¶ 120–124, 138–140, 145, 147, 218, Fig. 7.

SoundClear’s argument that Aoyama teaches away from replacing words based on distance fares no better. *See* PO Resp. 32; PO Sur-reply 18. 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 that Aoyama is limited to replacing words based on the “speech style of the user.” Ex. 1010 ¶ 121. But as we explain above, Aoyama is not so limiting because it also contemplates replacing words under other circumstances, such as depending on the “degree of composure” (e.g., calm, hurried, etc.) or depending on the situation at hand. *Id.* ¶¶ 138–140, 145, 147, 218. In our view, SoundClear attempts to read into Aoyama a teaching away where no such language exists. *See Dystar*, 464 F.3d at 1364.

c. Limitation 1[g]—“the output-content generating unit is further configured to . . . generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information”

In its Patent Owner Response, relying on the same arguments and evidence it presents for limitation 1[b][ii], SoundClear contends that Shin does not teach or suggest “generating a first and second output sentence when the voice has been classified as the first and second voice.” PO Resp. 37 (citing Ex. 1001; Ex. 2016 ¶¶ 107, 108). According to SoundClear, Amazon’s declarant, Dr. Stern, provides deposition testimony that improperly supplements the actual disclosure of Shin with his own assumptions to explain how Shin teaches or suggests a “second output sentence which includes all of the intention information and the content

information.” *Id.* (citing Ex. 1002 ¶ 106; Ex. 2017, 75:16–82:15).

SoundClear also argues that “Aoyama does not cure the deficiencies in Shin.” *Id.* (citing Ex. 2016 ¶ 108).

In its Reply, Amazon disagrees with SoundClear’s argument that Dr. Stern improperly supplemented the actual disclosure of Shin with his own assumptions. Pet. Reply 18. According to Amazon, both the Petition and Dr. Stern’s supporting testimony consistently show that Shin teaches or suggests “a second output sentence that includes both intention information and content information.” *Id.* (citing Pet. 28–29; Ex. 1002 ¶¶ 104–108; Ex. 2017, 82:2–6, 82:9–15, 130:12–132:2).

Based on the complete record, we agree with Amazon that Shin teaches or suggests, after classifying the voice as a second voice, generating a second output sentence that includes all the intention information and the content information, as required by limitation 1[g]. Pet. 28–29; Pet. Reply 18. As we explain above, 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.5.a. Shin’s Table 3 sets forth several levels of classification based on distance thresholds, one of which classifies the user’s voice as a second voice when the distance between the user and electronic device 101 is less than 1 meter. *See id.*

Shin further discloses that the output scheme associated with this classification level 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.” Ex. 1003 ¶ 87. This

means that, when classifying the user's voice as a second voice because the distance between the user and electronic device 101 is less than 1 meter, Shin discloses that electronic device 101 outputs the "amount of information of original content." *Id.* ¶ 86. As one example, when Shin's user 10 speaks "Let me know what time it is now," the electronic device generates the response "The current time is nine ten AM." *Id.* ¶ 37. Amazon argues—and we agree—that the word "time" in this response conveys all the intention information and the words "nine ten" in this response represent all the content information, both of which are derived from the user's statement "Let me know what time it is now." *See* Pet. 29; Ex. 1002 ¶ 107.

We do not agree with SoundClear's assertion that Dr. Stern's declaration and deposition testimony improperly supplement the actual disclosure of Shin with his own assumptions. *See* PO Resp. 37. SoundClear presents a naked assertion without adequately explaining how Dr. Stern's declaration and deposition testimony go above and beyond the actual disclosure in Shin that Amazon relies on to support its position that Shin teaches or suggests limitation 1[g]. *See* Ex. 1002 ¶¶ 106, 107; Ex. 2017, 75:16–82:15.

d. Whether Shin is required to disclose all the components as arranged in independent claim 1

In its Patent Owner Response, SoundClear contends that Amazon "focuses on the claimed functions without identifying what *component* allegedly accomplishes those functions." PO Resp. 38 (citing Ex. 2016 ¶ 109). SoundClear asserts that Amazon's obviousness ground based on the combined teachings of Shin and Aoyama "does not address the

details or arrangement of the components/steps of any claim of the '675 patent.” *Id.*

In its Reply, Amazon contends that its arguments and evidence with respect to the combined teachings of Shin and Aoyama that are set forth in the Petition “consistently identif[y] the components accomplishing the claimed functions, which are the same components disclosed in the '675 patent.” Pet. Reply 18–19 (citing Pet. 8–9 (processor/CPU), 11 (proximity sensor)).

Based on the complete record, we do not agree with SoundClear’s argument that Amazon fails to explain how Shin discloses all the components as arranged in independent claim 1. *See* PO Resp. 38. Amazon does not rely on Shin alone to teach or suggest the arrangement of components as required by independent claim 1, but rather Amazon relies on the combined teachings of Shin and Aoyama to account properly for this arrangement. 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 “voice acquiring unit,” “voice classifying unit,” “proximity sensor,” “intention analyzing unit,” “notification-information acquiring unit,” and “output-content generating unit” required by independent claim 1. *See* Pet. 7–29. To the extent Shin’s processor 120 does not teach or suggest the “intention analyzing unit” and “notification-information acquiring unit,” Amazon provides alternative arguments and evidence explaining how Aoyama’s meaning analysis unit 114 and response component generation unit 12 teach or suggest these two

claimed features. *See id.* at 18–20. Notably, when arguing that Shin does not disclose all the components as arranged in challenged claims, SoundClear does not identify, with particularity, a claimed component that is actually missing from Amazon’s proposed combination of Shin and Aoyama.

e. Summary

In summary, Amazon has demonstrated by a preponderance of the evidence that the subject matter of independent claims 1, 6, and 7 would have been obvious over the combined teachings of Shin and Aoyama.

6. Claims 2–4

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

2[a] the notification information includes type information indicating a type of the content information; and
2[b] the output-content generating unit is further configured to select, on generating the first output sentence, at least one word to be placed [sic] with the other word based on a word including in the content information and the type information that is associated with the content information.

Ex. 1001, 21:21–30 (with bracket annotations provided by Amazon in the Petition). Relying on the same arguments and evidence it presents for limitation 1[f], Amazon contends that Aoyama teaches or suggests limitation 2[a] because Aoyama’s information processing device 1 references replacement words stored in tables, such as Figure 7 that illustrates a diagram of an example of providing a response parameter based on “the speech style of the user.” Pet. 30–31 (citing Ex. 1010, Fig. 7). Amazon argues that Aoyama’s response parameter generation unit 16 and response generation unit 17 work together to provide notification information that

includes type information (e.g., the “name” field) that indicates a type of content information (e.g., “Taro Yamada”). *Id.* at 31 (citing Ex. 1010 ¶ 124; Ex. 1002 ¶ 112). Similarly, Amazon argues that Aoyama’s tables, such as the one illustrated in Figure 7, include type information because these tables correspond to the types of information the system looks up and replaces (e.g., the “name” field in Figure 7 is type information for content information such as “Taro Yamada” and “Mr. Yamada”). *Id.* at 31–32 (citing Ex. 1002 ¶ 113).

Once again, relying on the same arguments and evidence it presents for limitation 1[f], Amazon contends Aoyama teaches or suggests limitation 2[b] because it discloses replacing names, including replacing “Taro Yamada” with “Mr. Yamada,” according to the user’s speech style. Pet. 32–33 (citing Ex. 1010 ¶ 121). As one example, Amazon argues that Aoyama teaches selecting a word (e.g., “Taro”) and replacing it with another word (e.g., “Mr.”) based on a word (e.g., “Taro Yamada”) included in the content information and the type of information (e.g., the “name” field) associated with that content information. *Id.* at 33 (citing Ex. 1002 ¶ 116).

Claim 3 depends from dependent claim 2, and further recites the following:

the output-content generating unit is further configured to select, on generating the first output sentence, relationship information including information of a first word that is a predetermined word and a second word that is associated with the first word, and to replace a word included in the notification information with the second word when the word included in the notification information matches with the first word.

Ex. 1001, 21:31–40. Relying on the same arguments and evidence it presents for limitation 1[f], Amazon contends that Aoyama teaches or

suggests this limitation because Aoyama's information processing device 1 references replacement words from tables when generating an output sentence. Pet. 33. As one example with reference to Aoyama's Figure 7, Amazon argues that its information processing device 1 reads the table titled "personal name data," which stores a "name" field corresponding to the "formal name" of each person. *Id.* at 33–34 (citing Ex. 1010 ¶ 121, Fig. 7; Ex. 1002 ¶ 120). According to Amazon, Aoyama stores tables, such as the one illustrated in Figure 7, in advance of generating the output sentence so the words are predetermined. *Id.* at 34 (citing Ex. 1010 ¶¶ 119, 121, 122, 124, 198). Amazon further argues that Aoyama replaces "Taro Yamada" with "Mr. Yamada" to generate the response sentence "You have a meeting with Mr. Yamada in room A." *Id.* (citing Ex. 1010 ¶ 138).

Claim 4 depends from dependent claim 3, and further recites the following:

4[a] the relationship information further includes information of type information indicating a type of the first word; and

4[b] the output-content generating unit is further configured to select, on generating the first output sentence, a word included in the notification information as word to be replaced with another word when the type information of the word included in the notification information matches with the type information of the first word and the word included in the notification information matches with the first word.

Ex. 1001, 21:41–22:2 (with bracket annotations provided by Amazon in the Petition). Amazon contends that Aoyama teaches or suggests limitation 4[a] because its replacement word tables include type of information. Pet. 35. As one example, Amazon argues that Aoyama includes a table titled "personal name data" that includes "a 'name' corresponding to the 'formal

name’ [of] each person.” *Id.* at 35–36 (quoting Ex. 1010 ¶ 121) (citing Ex. 1010 ¶ 235, Figs. 7, 14). Relying on the testimony of Dr. Stern, Amazon asserts that the phrase “personal name” in the title “personal name data” indicates the type of words “Taro Yamada” and “Mr. Yamada” listed in the table.” *Id.* at 36 (citing Ex. 1002 ¶ 126).

Relying on the same arguments and evidence it presents for limitation 2[b], Amazon contends that Aoyama teaches or suggests limitation 4[b] because Aoyama’s information processing device 1 selects a word (e.g., “Taro Yamada”) in the notification information that is replaced by another word (e.g., “Mr. Yamada”) based on the relationship information set forth in the table illustrated in Figure 7. Pet. 36–37 (citing Ex. 1010 ¶¶ 121, 124). Relying on the testimony of Dr. Stern, Amazon argues that Aoyama requires that the type of information (e.g., “name”) for the word in the notification information (e.g., “Mr. Taro Yamada”) matches the type of information for the word in the relationship information (e.g. “name”) because the type of information determines which table information processing device 1 consults. *Id.* at 37 (citing Ex. 1002 ¶ 129). Amazon further argues that Aoyama’s information processing device 1 finds the entry in the personal name table for “Taro Yamada,” which requires matching the word in the notification information (e.g., “Taro Yamada”) with the word in the relationship information (e.g., Taro Yamada”). *Id.*

During trial, apart from relying on the same arguments it presents for independent claim 1, SoundClear does not address separately Amazon’s arguments and supporting evidence explaining how the combined teachings of Shin and Aoyama account for the additional limitations of dependent claims 2–4. *See* PO Resp. 38. For the same reasons we explain above with

respect to independent claim 1, we do not agree with SoundClear's arguments. *See supra* Sections II.B.5.a–e. We have reviewed Amazon's arguments and supporting evidence explaining how the combined teachings of Shin and Aoyama account for the additional limitations of dependent claims 2–4, and we agree with and adopt Amazon's analysis. *See* Pet. 30–37. Accordingly, SoundClear has demonstrated by a preponderance of the evidence that the subject matter of dependent claims 2–4 would have been obvious over the combined teachings of Shin and Aoyama.

7. Claim 5

Claim 5 depends from independent claim 1, and further recites “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” Ex. 1001, 22:3–7.⁷ Amazon contends that Shin teaches or suggests this limitation because electronic device 101 is capable of determining a user's speech volume and adjusting its output information based, at least in part, on the user's speech volume. Pet. 38 (citing Ex. 1003, Abstract, ¶¶ 37–39, 49, 50, 53, 54, 71, 76–78, 95, 137, claims 1–9, 15–17, 20; Ex. 1002 ¶ 132). Amazon argues that Shin's electronic device 101 “detects a range of speech volumes and rates.” *Id.* (citing Ex. 1003 ¶¶ 37, 38). Relying on the testimony of Dr. Stern, Amazon asserts that a person of ordinary skill in the art would have understood “Shin's disclosure to encompass receiving and processing as wide a range of speech as possible, including whispered speech.” *Id.* (citing Ex. 1002 ¶ 132).

⁷ *See supra* n.5.

Alternatively, Amazon argues that, even if Shin does not teach or suggest that electronic device 101 supports processing whispered speech, a person of ordinary skill in the art would have been motivated to configure electronic device 101 to receive and process whispered speech for several reasons. Pet. 38 (citing Ex. 1002 ¶¶ 133–137). As one example, Amazon argues that the prior art suggests doing so because Iyer discloses addressing privacy concerns by processing whispered speech. *Id.* (citing Ex. 1012, 3:61–64; Ex. 1002 ¶ 134). Amazon then provides four additional reasons as to why a person of ordinary skill in the art would have modified Shin’s electronic device 101 to receive and process whispered speech with a reasonable expectation of success. *See id.* at 39–40 (citing Ex. 1002 ¶¶ 135–137).

In its Patent Owner Response, SoundClear contends that Amazon “does not identify any express or inherent disclosure in Shin that allegedly meets the elements of [dependent] claim 5.” PO Resp. 36 (citing Ex. 2016 ¶¶ 105, 106). According to SoundClear, dependent claim 5 “requires more than the capacity to receive and process whispered speech”— “[i] requires a *determination* that ‘the voice spoken by the user is a whisper.’” *Id.* (citing Ex. 2016 ¶ 106). SoundClear asserts that such a “determination cannot be made if a whisper cannot be identified.” *Id.* SoundClear further argues that Amazon’s reliance on Iyer to teach or suggest the whispered speech required by dependent claim 5 is improper. *Id.* at 35. SoundClear acknowledges that a prior art reference may be used to establish the general background knowledge in the art, but it argues that, under these particular circumstances, Amazon improperly “provides Iyer to supply the missing limitation of determining a ‘whisper.’” *Id.*

In its Reply, Amazon contends that dependent claim 5 does not require determining whether the voice is a whisper, but instead “[it] requires that the device be configured to determine that ‘the voice’ is ‘the predetermined voice’ in circumstances where the user whispers to the device.” Pet. Reply 16–17 (citing Ex. 1001, claim 5; Ex. 1202 ¶ 19). Amazon asserts that “[n]othing in the claim requires the device to *determine* whether the received voice is a whisper.” *Id.* at 17. Amazon further argues that SoundClear mischaracterizes Amazon’s position with respect to dependent claim 5 set forth in the Petition because Amazon did not equate a whisper to be a range of speech volumes and rates. *Id.* According to Amazon, SoundClear fails to rebut its basic premise that a person of ordinary skill in the art would have understood Shin to encompass receiving and processing as wide a range of speech as possible, including whispered speech, or that it would have been obvious to configure Shin’s electronic device 101 to do so. *Id.* (citing Pet. 38–40).

Based on the complete record, we agree with SoundClear that Shin does not teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See* PO Resp. 33–36. As we explain in our claim construction analysis above, both the plain language of dependent claim 5 and the specification support requiring the voice classifying unit to determine whether the voice spoken by a user is a whisper. *See supra* II.A.2. The specification of the ’675 patent sets forth an explicit definition for the claim term “whisper” as “an unvoiced voice spoken without vibrating the vocal chords,” but it also includes a voice with a peak frequency equal to or lower than a threshold. *Id.* With this in mind, we address Amazon’s reliance on Shin to teach or suggest determining

whether the voice spoken by the user is a whisper, as required by dependent claim 5.

At best, Shin discloses that electronic device 101 receives and processes a wide range of speech volumes and rates. Ex. 1003 ¶¶ 37, 38. But Amazon does not direct us to—nor can we find—an explicit disclosure in Shin indicating that electronic device 101 determines whether the voice input received by audio input module 151 is whispered speech. Amazon’s reliance on the supporting testimony of Dr. Stern for this particular issue fares no better because he simply repeats the arguments and evidence presented in the Petition. *Compare* Pet. 38, *with* Ex. 1002 ¶ 132.

We recognize that Amazon attempts to remedy this deficiency in Shin by relying on Iyer to teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See* Pet. 38. There are two problems with this approach. First, we do not understand Amazon to rely on Iyer in a manner that serves to document the knowledge a person of ordinary skill in the art would bring to bear in reading Shin. *See Ariosa Diagnostics v. Verinata Health, Inc.*, 805 F.3d 1359, 1365 (Fed. Cir. 2015) (“Art can legitimately serve to document the knowledge that skilled artisans would bring to bear in reading the prior art identified as producing obviousness.” (citing *Randall Mfg. v. Rea*, 733 F.3d 1355, 1362–63 (Fed. Cir. 2013))). Instead, we understand Amazon to rely on Iyer to teach or suggest the limitation of determining whether the voice spoken by the user is a whisper that is actually missing from Shin. Second, even if we were to rely on Iyer to teach or suggest this limitation actually missing from Shin, Amazon did not raise an obviousness ground based on the combined teachings of Shin, Aoyama, and Iyer in the Petition. *See* Pet. 4 (disclosing a

table identifying the grounds of unpatentability raised by Amazon in the Petition, none of which include Iyer). It is well-settled that Amazon—as the Petitioner—is “master of its complaint.” *SAS Inst., Inc. v. Iancu*, 584 U.S. 357, 363 (2018). We are constrained by the arguments and evidence presented and developed in Amazon’s Petition, and we are not at liberty to “raise, address, and decide . . . theories never presented by [Amazon].” *In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1381 (Fed. Cir. 2016).

In summary, Amazon has not demonstrated by a preponderance of the evidence that the subject matter of dependent claim 5 would have been obvious over the combined teachings of Shin and Aoyama.

*C. Obviousness over Shin and Iwase*⁸

Amazon contends that claims 1, 2, and 5–7 of the ’675 patent are unpatentable under 35 U.S.C. § 103 as obvious over the combined teachings of Shin and Iwase. Pet. 41–46. Amazon contends that the combined teachings of Shin and Iwase account for the subject matter of each challenged claim, and 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

⁸ 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>.

Dr. Stern with the Petition and a Reply Declaration of Dr. Stern with the Reply. *See* Exs. 1002, 1202.

SoundClear contends that the proposed combination of Shin and Iwase does not teach or suggest limitation 1[f]. PO Resp. 39–42; PO Sur-reply 9–19. To support its positions, SoundClear submits the Declaration of Mr. McAlexander with the Patent Owner Response. *See* Ex. 2016.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1, 2, 6, and 7 of the '675 patent are unpatentable as obvious over the combined teachings of Shin and Iwase. We, however, determine that Amazon has not shown by a preponderance of evidence that claim 5 of the '675 patent is unpatentable as obvious over the combined teachings of Shin and Iwase. We begin our analysis by providing an overview of Iwase, and then we turn to the parties' contentions with respect to the challenged claims.

1. Iwase

Iwase generally relates to “an information processing device, and an information processing method.” Ex. 1017, 1:19–20. Iwase discloses information processing terminal 10, which “may be a mobile phone, a smart phone, a [PC], a tablet, or any of various wearable devices” (*id.* at 5:10–13), that “is capable of performing voice output of notification contents on the basis of a subject output position determined by . . . information processing server 30” (*id.* at 5:2–7, Fig. 2). Iwase discloses storing attribute information related to users, such as “age, distinction of sex, language, and hometown,” that “is used for shaping the notification contents.” *Id.* at 12:26–42. As one example, when a target user is a child, if the subject is “I recommend bringing an umbrella,” information processing terminal 10

works with information processing server 30 to reshape or adjust the subject by generating a response “You should take an umbrella.” *Id.* at 15:12–15.

2. Claims 1, 6, and 7

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. 76. For the most part, the same arguments and evidence Amazon provides to explain how the combined teachings of Shin and Iwase account for all the limitations of independent claim 1 equally apply to all the corresponding limitations of independent claims 6 and 7. *See id.* at 46. In addition, with the exception of limitation 1[f], Amazon’s arguments and evidence that it identifies above with respect to its obviousness ground based on the combined teachings of Shin and Aoyama equally apply to its obviousness ground based on the combined teachings of Shin and Iwase challenging these same claims. *See id.* at 41 (arguing “[Amazon] incorporates its discussion of Shin from [the obviousness grounds based on the combined teachings of Shin and Aoyama]”).

Limitation 1[f] recites “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word.” Ex. 1001, 21:9–17. Amazon contends that Iwase teaches or suggests this limitation because it discloses information processing terminal 10 that audibly presents notification information to users by adjusting the notification information based on user attributes, such as age and word preference. Pet. 41 (citing Ex. 1017, Abstract, 1:54–2:28, 3:26–29, 4:17–29, 5:2–7, 5:10–13, 5:63–6:3, 6:32–35, 12:26–42, 15:6–40;

Ex. 1002 ¶ 144). According to Amazon, Iwase's adjustment "may involve word replacement, such as replacing 'I recommend' in the 'I recommend bringing an umbrella' with 'You should,' yielding: 'You should take an umbrella.'" *Id.* at 42 (quoting Ex. 1017, 15:6–14).

Turning to rationale to combine and reasonable expectation of success, Amazon contends that there are several reasons why a person of ordinary skill in the art would have been motivated to configure Shin's processor 120 "to adjust the information provided to the user based on the user-to-device distance by incorporating Iwase's word replacement feature, for example to replace harder-to-understand words with simpler ones when the user is far away." Pet. 42 (citing Ex. 1002 ¶¶ 145–151). As one example, Amazon argues that Iwase provides an explicit teaching, suggestion, or motivation for configuring Shin's processor 120 in this way because Iwase discloses that its word replacement feature allows a user "to grasp notification contents more effectively." *Id.* (quoting Ex. 1017, 2:11–14) (citing Ex. 1017, 15:24–28, 21:66–22:8; Ex. 1002 ¶ 146). As another example, Amazon argues that configuring Shin's processor 120 in this way represents the simple addition of one known element (i.e., Iwase's word replacement feature) to another known element (i.e., Shin's distance-based output adjustment) to obtain predictable results (i.e., easier-to-understand output to users at greater distances). *Id.* (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶ 147). Amazon asserts that there would have been a reasonable expectation of success in incorporating Iwase's word replacement feature into Shin's processor 120 because "this configuration would have been trivial, involving straightforward changes to enhance the processor's approach to simplifying and abbreviating its output," "such an approach was

well known,” and “Iwase already contemplates incorporation of its word-replacement into a device like Shin’s, which analyzes the user-to-device distance and generates voice output.” *Id.* at 44 (citing Ex. 1002 ¶ 151; Ex. 1011).

In its Patent Owner Response, SoundClear contends “Iwase does not disclose replacing words based on voice classification.” PO Resp. 40 (citing Ex. 2016 ¶ 114). SoundClear argues that “[t]he word ‘replace’ does not appear in Iwase,” neither Amazon nor its declarant, Dr. Stern, explains where in Iwase a person of ordinary skill in the art would have understood there to be a word replacement, and Amazon does not provide any basis for an inherency position. *Id.* (citing Ex. 2016 ¶ 115). SoundClear further argues that, although Iwase discloses adjusting outputs based on the “*attribute information of a user*,” such as “age, distinction of sex, language, and hometown,” these operations do not involve any voice analysis, let alone voice classification. *Id.* (first quoting Ex. 1017, 15:4–10, then quoting Ex. 1017, 12:26–31) (citing Ex. 1017, 10:35–67, 15:41–45, Fig. 7; Ex. 2016 ¶ 116). Lastly, SoundClear asserts that a person ordinary skill in the art would have no reason to combine the teachings of Shin with those of Iwase because “Iwase is not concerned at all with distance” nor does it “improve any process that incorporates distance.” *Id.* at 41 (citing Pet. 42–44; Ex. 2016 ¶ 117).

In its Reply, Amazon contends SoundClear’s argument that Iwase does not disclose replacing words based on voice classification is irrelevant because Amazon relies on Shin—not Iwase—to teach or suggest the claimed voice classification. Pet. Reply 19 (citing Pet. 10–15; PO Resp. 39–40). Amazon also disagrees with SoundClear’s argument that the word “replace”

does not appear in Iwase because, according to Amazon, Iwase expressly discloses an example of replacing words where the words “I recommend” are replaced with “You should,” thereby providing output where at least one word “is replaced with another word,” as required by limitation 1[f]. *Id.* (citing PO Resp. 39–40; Ex. 1002 ¶ 144; Ex. 1017, 15:6–14). Lastly, Amazon argues that SoundClear’s rationale to combine argument is premised on the notion that Iwase is not concerned with distance, which is incorrect. *Id.* Amazon asserts that Iwase teaches determining the user-to-device distance and altering output based on that distance, and its declarant, Dr. Stern, explains why this disclosure in Iwase would have motivated a person of ordinary skill in the art to implement Iwase’s word replacement feature in Shin’s electronic device 101. *Id.* at 19–20 (citing Pet. 44; Ex. 1002 ¶ 150).

Based on the complete record, we agree with Amazon that the proposed combination of Shin and Iwase teaches or suggests, after classifying the voice is a first voice, generating a first output sentence that replaces one word with another word, as required by limitation 1[f]. Pet. 41–44; Pet. Reply 19–21. As we explain above, 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.5.a. More specifically, 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 1 to 2 meters. *See id.* Moreover, when classifying the user’s voice as a first voice because the distance between the user and

electronic device 101 is 1 to 2 meters, we find that Shin discloses that processor 120 outputs abbreviated or simplified content that may include omitting or replacing words. *See supra* Section II.B.5.b. However, we recognize that, even if we were to treat Shin’s abbreviation or simplification of content output to the user as limited to omitting words instead of replacing words, Iwase’s word replacement feature would fill this gap in Shin.

Iwase discloses an information processing system that includes, among other things, information processing terminal 10 and processing server 30. Ex. 1017, 4:51–64, Fig. 2. More specifically, Iwase discloses information processing terminal 10 “is capable of performing voice output of notification contents on the basis of a subject output position determined by . . . information processing server 30.” *Id.* at 5:2–7. Iwase discloses that information processing terminal 10 and processing server 30 work together to adjust the notification content provided to a user by using attribute information related to the user, such as “age, distinction of sex, language, and hometown.” *Id.* at 12:26–42. As one example of adjusting the notification content, Iwase disclose that, when a target user is a child, if the subject is “I recommend bringing an umbrella,” information processing terminal 10 works with information processing server 30 to adjust the subject by generating a response “You should take an umbrella.” *Id.* at 15:12–15.

Based on these cited disclosures, Amazon argues—and we agree—that Iwase teaches or suggests information processing terminal 10 that adjusts the notification content provided to the user by replacing words. As one example, we find that Iwase’s information processing terminal 10 is

capable of receiving the subject “I recommend bringing an umbrella” and replacing “I recommend” with “You should” so the resulting output to the user is “You should take an umbrella.”

Taking this teaching or suggestion from Iwase and applying it to Shin, Amazon argues—and we agree—that a person of ordinary skill in the art would have configured Shin’s processor 120 to classify the user’s voice based on distance thresholds between the user and electronic device 101 by incorporating Iwase’s word replacement feature to generate a first output sentence that replaces one word with another word, as required by limitation 1[f]. *See* Pet. 42. More specifically, after classifying the user’s voice as a first voice because the distance between the user and electronic device 101 is 1 to 2 meters, we find that Shin’s processor 120 is capable of outputting content to the user at a first abbreviation level by using Iwase’s word replacement feature, thereby generating an output sentence that replaces one word with another word (e.g., by generating an output sentence that takes the subject “I recommend bringing an umbrella” and replaces “I recommend” with “You should” so the resulting output to the user is “You should take an umbrella”).

We do not agree with SoundClear’s argument that the word “replace” does not appear Iwase and, therefore, it cannot teach or suggest replacing words as required by limitation 1[f]. *See* PO Resp. 40. When conducting an obviousness analysis, we must consider Iwase “not only for what it expressly teaches, but also for what it fairly suggests.” *Bradium*, 923 F.3d at 1049. As we explain above, Iwase’s information processing terminal 10 performs the word replacement required by limitation 1[f] because it adjusts the notification content output to a user by replacing words. Iwase’s word

replacement feature dictates the output scheme provided by Shin's electronic device 101 (e.g., by providing easier to understand output content to a user when the distance between the user and electronic device 101 is 1 to 2 meters).

Nor do we agree with SoundClear's argument that Iwase does not disclose operations based on voice classification. *See* PO Resp. 40. SoundClear's argument improperly attacks Iwase in isolation. *See In re Keller*, 642 F.2d at 426; *In re Young*, 927 F.2d at 591. As we explain above, Amazon's proposed obviousness ground relies on Shin—not Iwase—to teach or suggest the voice classification required by limitation 1[b][i]. *See* Pet. 12–15. Amazon turns to Shin's voice classification together with Iwase's word replacement feature to teach limitation 1[f]. *See id.* at 41–44.

Lastly, we do not agree with SoundClear's argument that a person of ordinary skill in the art would not have been motivated to combine the teachings of Shin with those of Iwase to arrive at limitation 1[f]. *See* PO Resp. 41. SoundClear's argument is premised on the notion that Iwase is not concerned with distance at all, nor does it improve any process that incorporates distance. This is incorrect. As Amazon correctly notes in its Reply, Iwase teaches or suggests determining the distance between the user and information processing terminal 10, which, in turn, is used to adjust the content output to the user. As one example, Iwase discloses that information processing server 30 includes image analysis unit 320 that “may recognize . . . a distance from the information processing terminal 10 and the user” by using, for example, “information collected by a depth sensor or the like.” Ex. 1017, 9:66–10:5, Fig. 5. As another example, Iwase discloses that information processing server 30 also includes situation estimation unit 330

that “is capable of detecting attention behavior on the basis that, for example, a user is approaching the information processing terminal 10.” *Id.* at 12:4–9, Fig. 5. In each of these examples, we find that Iwase teaches or suggests that the distance between the user and information processing terminal 10 is a factor in determining how to adjust the content output to the user.

In summary, Amazon has demonstrated by a preponderance of the evidence that the subject matter of independent claims 1, 6, and 7 would have been obvious over the combined teachings of Shin and Iwase.

3. *Claims 2 and 5*

Beginning with dependent claim 2, SoundClear does not address separately Amazon’s arguments and supporting evidence explaining how the combined teachings of Shin and Iwase account for the additional limitation of this dependent claim. *See* PO Resp. 39–42. We have reviewed Amazon’s arguments and supporting evidence explaining how the combined teachings of Shin and Iwase account for the additional limitations of dependent claim 2, and we agree with and adopt Amazon’s analysis. *See* Pet. 44–45. Accordingly, SoundClear has demonstrated by a preponderance of the evidence that the subject matter of dependent claim 2 would have been obvious over the combined teachings of Shin and Iwase.

Turning to dependent claim 5, Amazon relies on the same arguments and evidence it presents for the obviousness ground based on the combined teachings of Shin and Aoyama to support its obviousness ground based on the combined teachings of Shin and Iwase challenging this dependent claim. *See* Pet. 46. As we explain above, we agree with SoundClear that Shin does not teach or suggest determining whether the voice spoken by the user is a

whisper, as required by dependent claim 5. *See supra* Section II.B.7. In the Petition, Amazon does not rely on Iwase to cure this deficiency in Shin. *See* Pet. 46. Accordingly, SoundClear has not demonstrated by a preponderance of the evidence that the subject matter of dependent claim 5 would have been obvious over the combined teachings of Shin and Iwase.

D. Obviousness over Shimomura and Aoyama

Amazon contends that claims 1–7 of the '675 patent are unpatentable under 35 U.S.C. § 103 as obvious over the combined teachings of Shimomura and Aoyama. Pet. 46–60. Amazon contends that the combined teachings of Shimomura and Aoyama 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 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 proposed combination of Shimomura and Aoyama does not teach or suggest limitations 1[b][ii], 1[f], and 1[g]. PO Resp. 43–48, 51–52; PO Sur-reply 9–19. SoundClear also argues that Shimomura does not disclose all the components as arranged in independent claim 1. PO Resp. 52. 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 1–4, 6, and 7 of the '675 patent are unpatentable as obvious over the combined teachings of Shimomura and Aoyama. We, however, determine that Amazon has not

shown by a preponderance of evidence that claim 5 of the '675 patent is unpatentable as obvious over the combined teachings of Shimomura and Aoyama. We begin by 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. Shimomura, 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.” Ex. 1004 ¶ 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, and it does not 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.

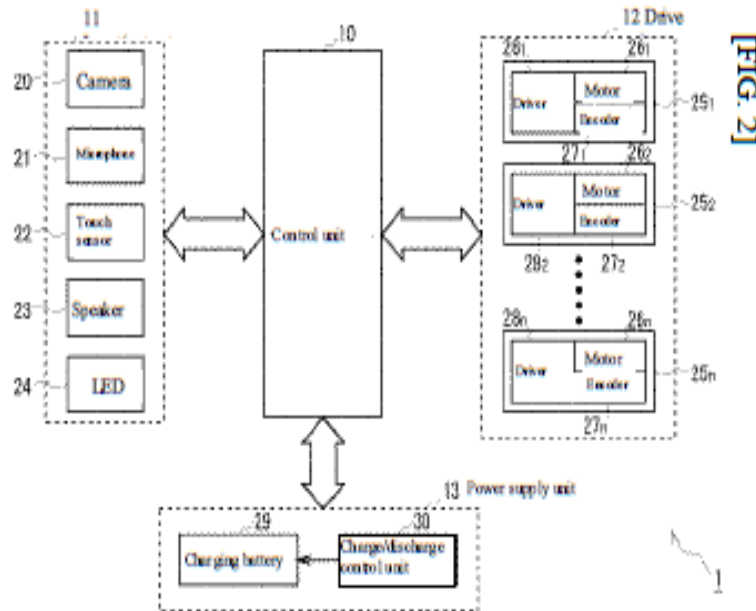


FIG. 2 Hardware constitution of the robot according to this embodiment (2)

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 corresponding 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. Claims 1, 6, and 7

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. 46–58, 76. For the most part, the same arguments and evidence Amazon provides to explain how the combined teachings of Shimomura and Aoyama account for all the limitations of independent claim 1 equally apply to the corresponding limitations of independent claims 6 and 7. *See* Pet. 60.

The preamble of independent claim 1 recites 1[pre]—“[a]n output-content control device.” Ex. 1001, 20:60. 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 receiving speech via microphones 21, performing speech recognition, and determining the “content to be spoken.” Pet. 46–47 (first quoting Ex. 1004, claim 1, then quoting Ex. 1004 ¶ 46) (citing Ex. 1004, Title, Abstract, ¶¶ 1, 30, Figs. 1, 4; Ex. 1002 ¶ 161). As one example, Shimomura discloses that, if a user asks about the weather, the device responds with “Today’s weather in Tokyo is sunny.” *Id.* at 47 (quoting Ex. 1004 ¶ 65) (citing Ex. 1002 ¶ 161).

Limitation 1[a] recites “a voice acquiring unit configured to acquire a voice spoken by a user.” Ex. 1001, 20:61–62. Amazon contends that Shimomura teaches or suggests this limitation because the device (i.e., robot

1) includes, among other things, control unit 10 and microphones 21. Pet. 48 (citing Ex. 1004, Fig. 2). Amazon argues that Shimomura's control unit 10, which includes a CPU, controls overall operation and other data processing of the device. *Id.* (citing Ex. 1004 ¶¶ 20, 27, 28, 32, 34, Fig. 3). Amazon further argues that Shimomura's control unit 10 includes speech recognition unit 42 that "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 the microphone 21." *Id.* at 48–49 (quoting Ex. 1004 ¶ 40) (citing Ex. 1004 ¶¶ 36–38, 42, Fig. 4).

Limitation 1[b] recites "a voice classifying unit configured to 1[b][i] calculate a distance between the user and the output-content control device by a proximity sensor to 1[b][ii] classify the voice into either a first voice or a second voice based on the calculated distance." Ex. 1001, 20:63–67. Amazon contends that Shimomura teaches or suggests limitations 1[b] and 1[b][i] because its device includes a "distance detection means for detecting the distance to the conversation partner user." Pet. 49 (quoting Ex. 1004 ¶ 93). Amazon argues that Shimomura's control unit 10 also includes image recognition unit 41 that may detect and calculate a distance to the user "based on the image signal S1 from the [charged coupled device] camera 20." *Id.* (quoting Ex. 1004 ¶ 39) (citing Ex. 1004 ¶¶ 39, 59, 70).

Amazon further contends that Shimomura teaches or suggests limitation 1[b][ii] because it discloses that control unit 10 also includes "speech synthesis unit [48]" and "speech character string transformation unit [51]." Pet. 49 (quoting Ex. 1004 ¶ 67) (citing Ex. 1004 ¶¶ 36, 37, 46, 58, Fig. 4). According to Amazon, these two units in Shimomura may

“transform the character string” of the content spoken by the device
“according to the distance to the conversation partner.” *Id.* at 50 (quoting Ex. 1004 ¶ 58) (citing Ex. 1004 ¶¶ 60–63, 67–70, 73–84, Abstract, claims 1–24). Amazon asserts that, prior to accomplishing this character string transformation, 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 greater than 350 centimeters (“cm”); and (2) a second voice at a distance less than 350 cm. *Id.* (citing Ex. 1002 ¶ 170). 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. The annotated version of Figure 7 illustrates a character string transformation table with the first voice at a distance greater 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.* (citing Ex. 1004 ¶¶ 60–64, Fig. 7; Ex. 1002 ¶¶ 169–171).

Limitation 1[c] recites “an intention analyzing unit configured to analyze the voice acquired by the voice acquiring unit to detect intention information indicating what kind of information is wished to be acquired by

the user.” Ex. 1001, 21:1–4. Limitation 1[d] recites “a notification-information acquiring unit configured to acquire notification information which includes content information as a content information to be notified to the user based on the intention information.” *Id.* at 21:5–8. Amazon contends that Shimomura teaches or suggests limitations 1[c] and 1[d] because control unit 10 is configured to analyze the voice acquired from speech recognition unit 42 by using the speech signal received by microphones 21 and, based on that speech, control unit 10 determines the content output to the user. Pet. 51 (citing Ex. 1004 ¶ 40; Ex. 1002 ¶ 173). Amazon argues that, because Shimomura describes the user as a conversation partner, this confirms that the device extracts user intent during conversation and responds by performing the requested processing. *Id.* (citing Ex. 1004 ¶¶ 3, 4, 10, 50, 54, 57–59, 67, 86, 87, 90, claims 1, 9, 17; Ex. 1002 ¶ 173).

Alternatively, Amazon contends that, even if Shimomura does not teach or suggest limitations 1[c] and 1[d], Aoyama teaches or suggests these two limitations. Pet. 53 (citing Ex. 1002 ¶ 179). That is, Amazon argues that Aoyama teaches or suggests limitation 1[c] because it discloses a meaning analysis unit 114 that uses natural language processing to determine the “meaning” of a user’s voice input and to specify a “task” to be executed. *Id.* at 18 (first quoting Ex. 1010 ¶ 72, then quoting Ex. 1010 ¶ 74) (citing Ex. 1010 ¶¶ 73, 75–84, 171–173, 230–232, Figs. 3, 8, 13). As one example, Amazon argues that Aoyama’s meaning analysis unit 114 receives a user’s voice input of “Check my schedule for tomorrow at 1 pm” and then detects the intended scheduled information. *Id.* (quoting Ex. 1010 ¶ 76) (citing Ex. 1010 ¶¶ 77–79, 170–173). Amazon also argues that Aoyama

teaches or suggests limitation 1[d] because it disclose response component generation unit 12 that “executes the task indicated by the information acquired from the meaning analysis unit 114 . . . and acquires information necessary to generate a response to the user.” *Id.* at 19 (quoting Ex. 1010 ¶ 127) (citing Ex. 1010 ¶¶ 128–130, 174–176, 233, Figs. 3, 8, 13).

For instance, in the example where the user’s voice input requests scheduling information, Aoyama discloses that response component generation unit 12 “acquires a search result” from a scheduling application. *Id.* (quoting Ex. 1010 ¶ 129).

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 Aoyama’s meaning analysis unit 114 and response component generation unit 12 in Shimomura’s robot for the same reasons a person of ordinary skill in the art would have been motivated to modify Shin’s electronic device 101. Pet. 53 (citing Ex. 1002 ¶ 179). That is, as one example, Amazon argues that the proposed combination would have reflected the addition of one known element (i.e., “Aoyama’s natural language processing and information retrieval”) to another known element (i.e., Shimomura’s device) to obtain predictable results (i.e., “identifying a user’s intent and providing information responsive to such intent”). *Id.* at 19–20 (citing *KSR*, 550 U.S. at 417). According to Amazon, there would have been a reasonable expectation of success in modifying Shimomura in this way because doing so would have been trivial, involving at most minor changes to Shimomura’s control unit 10 to implement Aoyama’s units. *See id.* at 20.

Limitation 1[e] recites “an output-content generating unit configured to generate an output sentence as sentence data to be output to the user based on the notification information.” Ex. 1001, 21:9–11. Relying on its arguments and evidence presented for limitations 1[b], 1[b][i], 1[b][ii], 1[c], and 1[d], Amazon contends that that Shimomura teaches or suggests limitation 1[e] because it discloses that control unit 10 generates content for output to the user. Pet. 53. For example, Amazon argues that, in Shimomura’s example where the user requests weather information, control unit 10 generates the output sentence “Today’s weather in Tokyo is sunny.” *Id.* (quoting Ex. 1004 ¶ 65) (citing Ex. 1004 ¶¶ 76, 78, 79, 92). Amazon further argues that Shimomura’s control unit 10 also includes behavior selection control unit 45 and output management unit 46 that work together to output a sentence as speech to the user. *Id.* (citing Ex. 1004 ¶¶ 36, 37, 46, 47, 49, 50, 63, 83, claims 6, 4, Figs. 4, 13).

Limitation 1[f] recites “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word.” Ex. 1001, 21:9–17. Amazon contends that the proposed combination of Shin and Aoyama teaches or suggests limitation 1[f] because Shimomura’s control unit 10 implements speech character string transformation unit 51 to insert a word like “ne” after each phrase so the output “can be clearly understood according to the physical distance to the conversation partner.” Pet. 54 (quoting Ex. 1004 ¶ 57) (citing Ex. 1004 ¶¶ 58–66, 85, 92, Figs. 7, 8). Amazon further argues that, when classifying a voice as a first voice (i.e., greater than 350 cm), Shimomura’s robot

generates a first output by transforming the original content to insert “ne” or another word after each phrase. *Id.* (citing Ex. 1004 ¶¶ 62, 65, 92, Fig. 7).

For example, Amazon argues that Shimomura discloses “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 55 (quoting Ex. 1004 ¶ 65) (citing Ex. 1002 ¶ 184) (alteration in original).

Amazon turns to Aoyama to more clearly teach or suggest replacing words to generate Shimomura’s output content. Pet. 55 (citing Ex. 1002 ¶¶ 183–185). That is, Amazon argues that Aoyama’s device (e.g., a smartphone) recognizes and analyzes a user’s speech, executes processing based on the voice input, and audibly presents the result as a sentence. *Id.* at 23 (citing Ex. 1010 ¶¶ 42, 51, 127, 161, 241–243, 253, Fig. 8). Amazon asserts that Aoyama’s device tailors output by replacing words based on the user or situation, such as by replacing a user’s formal name with an informal name. *Id.* at 23–24 (citing Ex. 1010 ¶¶ 120–124, 138–140, 177, 218, 234–236, Figs. 7, 14; Ex. 1002 ¶ 94).

Turning to rationale to combine and reasonable expectation of success, Amazon contends a person of ordinary skill in the art would have been motivated to adjust Shimomura’s output content using Aoyama’s word replacement feature for several reasons. Pet. 25 (citing Ex. 1002 ¶¶ 186–191). As one example, Amazon argues that Shimomura suggests doing so because it discloses that providing clear enunciation, particularly at distance, is advantageous. *Id.* at 55 (citing Ex. 1004 ¶¶ 4–17; Ex. 1002 ¶ 187). According to Amazon, this proposed combination would result in replacing harder to understand formal names with easier to understand informal

names, thereby motivating a person of ordinary skill in the art to use Aoyama's word replacement feature to advance Shimomura's goal of providing clear enunciation. *Id.* As another example, Amazon argues that the proposed combination merely represents the simple addition of one known element (i.e., "Aoyama's word replacement" feature) to another known element (i.e., "Shimomura's distance-based content output adjustments") to obtain predictable results (i.e., adjusting output content "based on distance using word replacement). *Id.* at 55–56 (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶ 188). According to Amazon, there would have been a reasonable expectation of success in making this proposed combination because (1) "this configuration would have been trivial, involving straightforward changes to enhances the control unit [10's] approach to adjusting output"; and (2) "such approach was well known." *Id.* at 56 (citing Ex. 1002 ¶ 190).

Limitation 1[g] recites "wherein the output-content generating unit is further configured . . . to generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information." Ex. 1001, 21:11–20. Amazon contends that Shimomura teaches or suggests this limitation because it discloses that speech characters 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. 57 (quoting Ex. 1004 ¶ 62) (citing Ex. 1004 ¶¶ 57, 58, 65, 85, 92) (alterations in original). According to Amazon, when Shimomura's device classifies the voice as a second voice (e.g., less than 350 cm), it generates an output sentence comprising the original content, without adjustment. *Id.* (citing Ex. 1004

¶¶ 62, 92, Fig. 7). For example, Amazon argues that, when classifying the voice as a second voice, the output sentence may be “Today’s weather in Tokyo is sunny.” *Id.* at 57–58 (quoting Ex. 1004 ¶ 65). In this weather example, Amazon asserts that the word “weather” conveys all the intent information derived from a user’s statement “Let me know today’s weather,” and the words “Tokyo” and “sunny” represent all the content information derived from that user statement. *Id.* (citing Ex. 1002 ¶ 194).

Notwithstanding SoundClear’s arguments to the contrary, Amazon has shown by a preponderance of the evidence that the combined teachings of Shimomura and Aoyama account for all the limitations of independent claims 1, 6, and 7. During trial, SoundClear contends that the proposed combination of Shimomura and Aoyama does not teach or suggest limitations 1[b][ii], 1[f], and 1[g]. PO Resp. 43–48, 51–52; PO Sur-reply 9–19. SoundClear also argues that Shimomura does not disclose all the components as arranged in independent claim 1. PO Resp. 52. We begin our analysis by addressing SoundClear’s arguments directed to limitations 1[b][ii], 1[f], and 1[g], and then we address SoundClear’s arguments directed to whether Shimomura discloses the claimed components as arranged in independent claim 1.

a. Limitation 1[b][ii]—“ a voice classifying unit configured to . . . classify the voice into either a first voice or a second voice based on the calculated distance”

SoundClear contends that Shimomura “does not disclose any manner of classifying voices.” PO Resp. 43 (citing Ex. 2016 ¶ 123). According to SoundClear, the only basis Amazon relies on to teach or suggest voice classification 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 43–44. SoundClear argues that Amazon also does not provide a reasoned analysis as to why a person of ordinary would equate measurement of distance with either “classification” or “voice.” *Id.* at 44 (citing Ex. 2016 ¶ 124). SoundClear argues that, to the extent Shimomura discloses the claimed classification, it only discloses classes of distances because Shimomura’s Figure 7 merely illustrates “a calculation of distance and nothing else.” *Id.* (citing Pet. 5; Ex. 1004, Fig. 7; Ex. 2016 ¶¶ 125, 126); *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 45 (citing Ex. 1004 ¶ 4; Ex. 2016 ¶ 128). SoundClear asserts that, “like Shin, Shimomura directly correlates the output content and volume based on distance without performing the intermediate step of classifying the voice.” *Id.* (citing Ex. 1004 ¶¶ 58, 67–70; Ex. 2016 ¶ 129).

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][ii] requires. Pet. Reply 21–22 (citing Pet. 48–50). Amazon argues that, contrary to SoundClear’s assertions, Shimomura’s Figure 7 discloses more than calculation of distance—this figure discloses classification of voice input based on whether a user’s distance exceeds a threshold of 350 cm. *Id.* (citing PO Resp. 44; Ex. 1201, 21:15–8). Amazon argues that SoundClear ignores that Shimomura’s Figure 7 describes “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.* (citing Ex. 1004, Fig. 7; Ex. 1201, 21:15–22; Ex. 1202 ¶ 21).

In its Sur-reply, SoundClear contends that Shimomura does not disclose classifying a “voice” because distance thresholds alone “do not . . . delineate one class of voice from another.” PO Sur-reply 9. According to SoundClear, Amazon has not demonstrated that Shimomura’s distance thresholds are capable of classifying voices, nor has Amazon shown that a person of ordinary skill in the art would have reasonably understood Shimomura to teach or suggest actually doing so. *Id.* Bottom line, SoundClear asserts that “[t]he Petition is devoid of any reasoned analysis” that Shimomura “*actually* classifies the voice.” *Id.* at 9–10. According to SoundClear, Dr. Stern provides deposition testimony confirming that Shimomura “merely discloses measuring variables (e.g., distance) and modifying output content based on the measurement,” but never discloses voice classification. *Id.* at 10–13 (citing PO Resp. 12–13; Ex. 2017, 57:6–13, 57:18–58:5, 58:6–12, 60:13–61:4, 95:7–17, 119:13–19; Ex. 2021, 21:8–22:16, 33:5–34:15, 62:22–63:8, 64:5–65:2, 65:6–12, 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 Shimomura to disclose voice classification. *See id.* at 13. At best, SoundClear asserts that Shimomura discloses classifying distances rather than classifying voices, as required by the challenged claims. *Id.* at 14–15 (citing Pet. Reply 11; Ex. 2017, 48:3–7, 49:7–15, 117:1–21, 118:2–13, 119:6–22).

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][ii]. Pet. 49–50; Pet. Reply 21–22.

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 greater 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 greater 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. 49–50 (citing Ex. 1004, Fig. 7 (with Amazon’s annotations of Figure 7 including the claimed “first voice” and “second voice” (in red))); Pet. Reply 21–22. 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][ii].

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 to the user and, therefore, it does not teach or suggest voice classification. *See* PO Resp. 43–44. When conducting an obviousness analysis, we must consider Shimomura “not only for what it expressly teaches, but also for what it fairly suggests.” *Bradium*, 923 F.3d at 1049. As we explain above, Shimomura's control unit 10 implements software that teaches or suggests performing the voice classification required by limitation 1[b][ii] 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. 44–45; PO Sur-reply 10–13. 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][ii]. *See* Pet. 49–50; Pet. Reply 21–22. 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 is greater 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. 45 (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.

b. Limitation 1[f]—"the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence in which at least one word selected among words included in the content information of the notification information is replaced with another word"

In its Patent Owner Response, relying on the same arguments it presents to rebut the obviousness ground based on the combined teachings of Shin and Aoyama, SoundClear contends that Amazon's proposed combination of Shimomura and Aoyama does not teach or suggest limitation

1[f]. PO Resp. 46 (citing Ex. 2016 ¶ 130). SoundClear also reiterates its argument that Aoyama’s word choices based on the “speech style of the user” and its change in “amount of information” (“similar to Shin’s content abbreviation”) is based on the user’s “degree of composure,” but these two features “involve separate and distinct processes and metrics.” *Id.* (first quoting Ex. 1010 ¶ 121, then quoting Ex. 1010 ¶ 145) (citing Ex. 1010 ¶¶ 122–124, 146, 147; Ex. 2016 ¶ 131) *see also id.* at 47–48 (arguing the differences between Aoyama’s disclosure of inputting words based on the “speech style of the user” versus changing the “amount of information output as a response” based on a user’s “degree of composure”).

SoundClear argues that a person of ordinary skill in the art would not have modified Shimomura to include Aoyama’s word replacement in Shimomura’s output content based on user-to-device distance. *Id.* at 46–47. Instead, SoundClear asserts that a person of ordinary skill in the art would have understood that Aoyama teaches away from replacing words based on user-to-device distance because “Aoyama teaches that words are *only* to be chosen in response to ‘speaking style’” and “[i]t would have led to a response with which ‘the user may feel uncomfortable.’” *Id.* (quoting Ex. 1010 ¶ 49) (citing Ex. 2016 ¶ 131).

In its Reply, Amazon contends that SoundClear’s arguments that a person of ordinary skill in the art would not have incorporated Aoyama’s word replacement feature in Shimomura “are identical to the arguments made with respect to Shin” and, therefore, they fail for the same reasons. Pet. Reply 23.

Based on the complete record, we agree with Amazon that the proposed combination of Shimomura and Aoyama teaches or suggests, after

classifying the voice as a first voice, generating a first output sentence that replaces one word with another word, as required by limitation 1[f].

Pet. 54–57; Pet. Reply 23. As we explain above, 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.D.2.a. More specifically, 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 greater than 350 cm. *See id.*

Shimomura further discloses that the output scheme associated with this classification level involves control unit 10 using speech character string transformation unit 51 to insert a word like “ne” after each phrase so the output “can be clearly understood according to the physical distance to the conversation partner.” Ex. 1004 ¶ 57. This means that, when classifying the user’s voice as a first voice because the distance between the conversation partner user and the robot is greater than 350 cm, Shimomura’s control unit 10 transforms the original content by adding character strings to the output content. *Id.* ¶¶ 57, 62, Fig. 7. As one example, 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.* ¶ 65.

Based on the complete record, we agree with Amazon that Shimomura teaches or suggests generating a first output sentence that transforms the original content based on classifying a user’s voice as a first voice. Amazon, however, turns to Aoyama to teach or suggest the word

replacement feature missing from Shimomura. Pet. 54–57. As we explain above, Aoyama teaches or suggests that information processing device 1 tailors the content output to the user by replacing words. *See supra* Section II.B.5.b. As one example pertaining to a user’s name, we find that Aoyama’s information processing device 1 is capable of receiving the name “Taro Yamada” and then replacing “Taro” with “Mr.” so the resulting name output to the user is “Mr. Yamada.” *See id.*

Taking this teaching or suggestion from Aoyama and applying it to Shimomura, Amazon argues—and we agree—that a person of ordinary skill in the art would have configured Shimomura’s control unit 10 that classifies the user’s voice based on distance thresholds between the conversation partner user and the robot by incorporating Aoyama’s word replacement feature to generate an output sentence that replaces one word with another word, as required by limitation 1[f]. More specifically, after classifying the user’s voice as a first voice because the distance between the user and the robot is greater than 350 cm, we find that Shimomura’s control unit 10 is capable of outputting a sentence to the user by transforming the original content using Aoyama’s word replacement feature, thereby generating an output sentence that replaces one word with another word (e.g., by generating an output sentence that takes the name “Taro Yamada” and replaces “Taro” with “Mr.” so the resulting name output to the user is “Mr. Yamada”). In our view, configuring Shimomura’s control unit 10 in this way compliments Shimomura’s goal of providing a robot with clear enunciation, particularly at greater distances. *See, e.g.*, 1004 ¶ 17 (disclosing that “a robot device is provided with a speech style change means for changing the speech style when interacting with a user as

necessary depending on the distance to the user, so that the robot device can always speak in a style that is easy for the user to hear, allowing for smooth interaction with the user”).

We do not agree with SoundClear’s argument that a person of ordinary skill in the art would not have been motivated to combine the teachings of Shimomura with those of Aoyama to arrive at limitation 1[f]. *See* PO Resp. 46–47. To support its assertion that Aoyama teaches away from replacing words based on distance, SoundClear direct us to Aoyama’s statement that “the user may feel uncomfortable.” Ex. 1010 ¶ 49. Once again, SoundClear fails to recognize that this statement in Aoyama refers to a user receiving a response from information processing device 1 in a speaking style different from the user’s preferred speaking style (e.g., receiving a response of “15:00” rather than a response of “3 pm”). *Id.* As we explain, Aoyama provides a solution to this problem by providing several different options for replacing words, including based on “the speech style of use,” based on the “degree of composure” (e.g., calm, hurried, etc.), or depending on the situation at hand. *Id.* ¶¶ 120–124, 138–140, 145, 147, 218, Fig. 7.

SoundClear’s argument that Aoyama teaches away from replacing words based on distance fares no better. *See* PO Resp. 47. SoundClear’s teaching away argument is, once again, predicated on the notion that that Aoyama is limited to replacing words based on the “speech style of the user.” Ex. 1010 ¶ 121. But as we explain above, Aoyama is not so limiting because it also contemplates replacing words under other circumstances, such as depending on the “degree of composure” (e.g., calm, hurried, etc.) or depending on the situation at hand. *Id.* ¶¶ 138–140, 145, 147, 218. In our

view, SoundClear attempts to read into Aoyama a teaching away where no such language exists. *See Dystar*, 464 F.3d at 1364.

c. Limitation 1[g]—“the output-content generating unit is further configured to . . . generate, when the voice is determined to be the second voice, a second output sentence which includes all of the intention information and the content information”

In its Patent Owner Response, relying on the same arguments and evidence it presents for limitation 1[b][ii], SoundClear contends that Shimomura does not teach or suggest “generating a first and second output sentence when the voice has been classified as the first and second voice.” PO Resp. 51 (citing Ex. 2016 ¶¶ 139, 140). According to SoundClear, Amazon’s declarant, Dr. Stern, provides deposition testimony that improperly supplements the actual disclosure of Shimomura with his own assumptions to explain how Shimomura teaches or suggests a “second output sentence which includes all of the intention information and the content information.” *Id.* at 51–52 (citing Ex. 1002 ¶ 193; Ex. 2017, 124:1–125:19, 126:4–9; Ex. 2016 ¶ 141).

In its Reply, Amazon disagrees with SoundClear’s argument that Shimomura does not teach or suggest the second output sentence that includes both intention information and content information, as required by limitation 1[g]. Pet. Reply 23. According to Amazon, Shimomura provides an example of output sentence that satisfies limitation 1[g]: “Today’s weather in Tokyo is sunny.” *Id.* (quoting Ex. 1004 ¶ 65) (citing Pet. 57–58). According to Amazon, this sentence in Shimomura “conveys all the intention information (‘weather’) and content information (‘Tokyo’ and ‘sunny’).” *Id.* (citing Ex. 1002 ¶ 194).

Based on the complete record, we agree with Amazon that Shimomura teaches or suggests that, after classifying the voice as a second voice, generating a second output sentence that includes all the intention information and the content information, as required by limitation 1[g]. Pet. 57–58; Pet. Reply 23. As we explain above, 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.D.2.a. More specifically, Shimomura’s Figure 7 sets forth two levels of classification based on distance thresholds, one of which classifies the user’s voice as a second voice when the distance between the user and the robot is less than 350 cm. *See id.*

Shimomura further discloses that the output scheme associated with this classification level involves control unit 10 using speech character string transformation unit 51 to determine that “the character string should not be transformed if the distance to the conversation partner . . . is less than 350 [cm].” Ex. 1004 ¶ 62. This means that, when classifying the user’s voice as a second voice because the distance between the conversation partner user and the robot is less than 350 cm, Shimomura’s robot generates an output sentence comprising the original content, without adjustment. *Id.* ¶¶ 57, 62, 63 Fig. 7. As one example, when the conversation partner user requests information regarding the weather, Shimomura’s robot generates the response “Today’s weather in Tokyo is Sunny.” *Id.* ¶ 65. Amazon argues—and we agree—that the word “weather” in this response conveys all the intention information and the words “Tokyo” and “sunny” in this response represent all the content information, both of which are derived from the

user's statement "Let me know today's weather." *See* Pet. 57–58; Ex. 1002 ¶ 194.

We do not agree with SoundClear's assertion that Dr. Stern's declaration and deposition testimony improperly supplement the actual disclosure of Shimomura with his own assumptions. *See* PO Resp. 51–52. SoundClear presents a naked assertion without adequately explaining how Dr. Stern's declaration and deposition testimony go above and beyond the actual disclosure in Shimomura that Amazon relies on to support its position that Shimomura teaches or suggests limitation 1[g]. *See* Ex. 1002 ¶¶ 193, 194; Ex. 2017, 124:1–125:19, 126:4–9.

d. Whether Shimomura is required to disclose all the components as arranged in independent claim 1

In its Patent Owner Response, SoundClear contends that Amazon "focuses on the claimed functions without identifying what *component* allegedly accomplishes those functions." PO Resp. 52 (citing Ex. 2016 ¶ 142). SoundClear asserts that Amazon's obviousness ground based on the combined teachings of Shimomura and Aoyama "does not address the details or arrangement of the components/steps of any claim of the '675 patent." *Id.*

In its Reply, Amazon does not agree with SoundClear's argument that it fails to identify in the Petition the components that accomplish each of the claimed functions. Pet. Reply 24 (citing Pet. 48–49 (processor/CPU), 49 (camera or other distance sensor)).

Based on the complete record, we do not agree with SoundClear's argument that Amazon fails to explain how Shimomura discloses all the components as arranged in independent claim 1. *See* PO Resp. 52. Amazon

does not rely on Shimomura alone to teach or suggest the arrangement of components required by independent claim 1, but rather Amazon relies on the combined teachings of Shimomura and Aoyama to account properly for this arrangement. For example, consistent with Amazon's claim mapping of independent claim 1, Shimomura's Figure 2 illustrates that the hardware of the robot includes, among other things, control unit 10, cameras 20, microphones 21, and speakers 23. Shimomura's Figure 4 illustrates that behavior control system 40 of the robot includes, among other things, image recognition unit 41, speech recognition unit 42, speech synthesis unit 48, and speech character string transformation unit 51. In the Petition, Amazon explains how each of the components in Shimomura are arranged in the same manner as the "voice acquiring unit," "voice classifying unit," "proximity sensor," "intention analyzing unit," "notification-information acquiring unit," and "output-content generating unit" required by independent claim 1. *See* Pet. 48–60. To the extent Shimomura's control unit 10 does not teach or suggest the "intention analyzing unit" and "notification-information acquiring unit," Amazon provides alternative arguments and evidence explaining how Aoyama's meaning analysis unit 114 and response component generation unit 12 teach or suggest these two claimed features. *See id.* at 51–53. Notably, when arguing that Shimomura does not disclose all the components as arranged in independent claim 1, SoundClear does not identify, with particularity, a claimed component that is actually missing from Amazon's proposed combination of Shimomura and Aoyama.

e. Summary

In summary, Amazon has demonstrated by a preponderance of the evidence that the subject matter of independent claims 1, 6, and 7 would have been obvious over the combined teachings of Shimomura and Aoyama.

3. Claims 2–4

Amazon contends that Aoyama teaches or suggests the additional limitations of dependent claims 2–4 for the same reasons it identifies above in its obviousness ground based on the combined teachings of Shin and Aoyama. Pet. 58. During trial, SoundClear does not address separately Amazon’s arguments and supporting evidence explaining how the combined teachings of Shimomura and Aoyama account for the additional limitations of dependent claims 2–4. *See* PO Resp. 52. We have reviewed Amazon’s arguments and supporting evidence explaining how the combined teachings of Shimomura and Aoyama account for the additional limitations of these dependent claims, and we agree with and adopt Amazon’s analysis. *See* Pet. 30–37, 58. Accordingly, SoundClear has demonstrated by a preponderance of the evidence that the subject matter of dependent claims 2–4 would have been obvious over the combined teachings of Shimomura and Aoyama.

4. Claim 5

Claim 5 depends from independent claim 1, and further recites “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” Ex. 1001, 22:3–7.⁹ Relying on its arguments and evidence presented for

⁹ *See supra* n.5.

limitation 1[a], Amazon contends that Shimomura renders this limitation obvious because its device receives and processes speech input from a conversation partner user. Pet. 58. Relying on the testimony of Dr. Stern, Amazon asserts that a whisper is a form of speech. *Id.* (citing Ex. 1002 ¶ 198).

Alternatively, Amazon argues that, even if Shimomura does not teach or suggest receiving and processing whispered speech, a person of ordinary skill in the art would have been motivated to configure its device to receive and process whispered speech for several reasons. Pet. 58 (citing Ex. 1002 ¶¶ 199–203). As one example, Amazon argues that the prior art suggests doing so because Iyer discloses addressing privacy concerns by processing whispered speech. *Id.* at 59 (citing Ex. 1012, 3:61–64; Ex. 1002 ¶ 200). Amazon then provides three additional reasons as to why a person of ordinary skill in the art would have modified Shimomura’s device to receive and process whispered speech with a reasonable expectation of success. *See id.* (citing Ex. 1002 ¶¶ 201, 202; Exs. 1012, 1016).

In its Patent Owner Response, SoundClear contends that “there is no stated basis upon which one would conclude that Shimomura discloses any manner of whisper determination.” PO Resp. 48. More specifically, SoundClear argues that Amazon “does not identify any express or inherent disclosure in Shimomura that allegedly meets the elements of [dependent] claim 5.” *Id.* at 49 (citing Ex. 2016 ¶¶ 136, 137). According to SoundClear, dependent claim 5 “requires more than the capacity to receive and process whispered speech”— “[i]t requires a determination that ‘the voice spoken by the user is a whisper.’” *Id.* at 50 (citing Ex. 2016 ¶ 137). SoundClear asserts that such a “determination cannot be made if a whisper cannot be

identified.” *Id.* (citing Ex. 1001, 9:58–10:8). SoundClear further argues that Dr. Stern’s supporting testimony ignores the plain language of dependent claim 5. *Id.* at 48–49 (citing Ex. 2017, 93:15–18); *see also id.* at 51 (arguing the same).

In its Reply, Amazon contends that SoundClear does not dispute that Shimomura “receives and processes speech input,” “whisper is a form of speech,” and “Shimomura discloses classifying a whispered voice.” Pet. Reply 23 (quoting Pet. 58–60) (citing PO Resp. 48–51). Amazon also argues that SoundClear does not dispute that a person of ordinary skill in the art would have been motivated to modify Shimomura’s robot to receive and process whispered speech. *Id.*

Based on the complete record, we agree with SoundClear that Shimomura does not teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See* PO Resp. 48–51. As we explain in our claim construction analysis above, both the plain language of dependent claim 5 and the specification support requiring the voice classifying unit to determine whether the voice spoken by a user is a whisper. *See supra* II.A.2. The specification of the ’675 patent sets forth an explicit definition for the claim term “whisper” as “an unvoiced voice spoken without vibrating the vocal chords,” but it also includes a voice with a peak frequency equal to or lower than a threshold. *Id.* With this in mind, we address Amazon’s reliance on Shimomura to teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5.

At best, Shimomura discloses that the robot receives and processes “the speech of the user.” Ex. 1004 ¶ 46. But Amazon does not direct us

to—nor can we find—an explicit disclosure in Shimomura indicating that the robot determines whether the speech input signal received by microphones 21 is whispered speech. Amazon’s reliance on the supporting testimony of Dr. Stern for this particular issue fares no better because he simply repeats the arguments and evidence presented in the Petition.

Compare Pet. 58, *with* Ex. 1002 ¶ 198.

We recognize that Amazon attempts to remedy this deficiency in Shimomura by relying on Iyer to teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See* Pet. 59. There are two problems with this approach. First, we do not understand Amazon to rely on Iyer in a manner that serves to document the knowledge a person of ordinary skill in the art would bring to bear in reading Shimomura. *See Ariosa*, 805 F.3d at 1365. Instead, we understand Amazon to rely on Iyer to teach or suggest the limitation of determining whether the voice spoken by the user is a whisper that is actually missing from Shimomura. Second, even if we were to rely on Iyer to teach or suggest this limitation actually missing from Shimomura, Amazon did not raise an obviousness ground based on the combined teachings of Shimomura, Aoyama, and Iyer in the Petition. *See* Pet. 4 (disclosing a table identifying the grounds of unpatentability raised by Amazon in the Petition, none of which include Iyer). It is well-settled that Amazon—as the Petitioner—is “master of its complaint.” *SAS*, 584 U.S. at 363. We are constrained by the arguments and evidence presented and developed in Amazon’s Petition, and we are not at liberty to “raise, address, and decide . . . theories never presented by [Amazon].” *In re Magnum Oil Tools*, 829 F.3d at 1381.

In summary, Amazon has not demonstrated by a preponderance of the evidence that the subject matter of dependent claim 5 would have been obvious over the combined teachings of Shimomura and Aoyama.

E. Obviousness over Shimomura and Iwase

Amazon contends that claims 1, 2, and 5–7 of the '675 patent are unpatentable under 35 U.S.C. § 103 as obvious over the combined teachings of Shimomura and Iwase. Pet. 60–62. Amazon contends that the combined teachings of Shimomura and Iwase 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 obviousness ground based on the combined teachings of Shin and Iwase, SoundClear asserts that the combined teachings of Shimomura and Iwase do not teach or suggest limitation 1[f]. PO Resp. 39–42; PO Sur-reply 9–19. To support its positions, SoundClear submits the Declaration of Mr. McAlexander with the Patent Owner Response. *See* Ex. 2016.

Based on the complete record, we determine that Amazon has shown by a preponderance of the evidence that claims 1, 2, 6, and 7 of the '675 patent are unpatentable as obvious over the combined teachings of Shimomura and Iwase. We, however, determine that Amazon has not shown by a preponderance of evidence that claim 5 of the '675 patent is unpatentable as obvious over the combined teachings of Shimomura and

Iwase. We begin our analysis by addressing the parties' contentions with respect to the challenged claims.

1. Claims 1, 6, and 7

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. 76. For the most part, the same arguments and evidence Amazon provides to explain how the combined teachings of Shimomura and Iwase account for all the limitations of independent claim 1 equally apply to all the corresponding limitations of independent claims 6 and 7. *See id.* at 62. In addition, with the exception of limitation 1[f], Amazon's arguments and evidence that it identifies above with respect to its obviousness ground based on the combined teachings of Shimomura and Aoyama equally apply to its obviousness ground based on the combined teachings of Shimomura and Iwase challenging these same claims. *See id.* at 60 (arguing "[Amazon] incorporates its discussion of Shimomura [from the obviousness ground based on the combined teachings of Shimomura and Aoyama]"). With respect limitation 1[f], Amazon's arguments and evidence that it identifies above with respect to its obviousness ground based on the combined teachings of Shin and Iwase equally apply to its obviousness ground based on the combined teachings of Shimomura and Iwase challenging this limitation. *See id.* (arguing "[Amazon] incorporates its discussion of . . . Iwase [from the obviousness ground based on the combined teachings of Shin and Iwase]").

Turning to rationale to combine and reasonable expectation of success, Amazon contends that there are several reasons why a person of ordinary skill in the art would have been motivated to configure

Shimomura’s control unit 10 “to adjust the output based on the user-to-device distance by incorporating Iwase’s word replacement feature, for example to replace harder-to-understand words with simpler ones when the user is far away.” Pet. 61 (citing Ex. 1002 ¶¶ 208–212). As one example, Amazon argues that Iwase provides an explicit teaching, suggestion, or motivation for configuring Shin’s processor 120 in this way because Iwase discloses that its word replacement feature allows a user “to grasp notification contents more effectively.” *Id.* (quoting Ex. 1017, 2:11–14) (citing Ex. 1017, 15:24–28, 21:66–22:8; Ex. 1002 ¶ 146). As another example, Amazon argues that configuring Shimomura’s control unit 10 in this way represents the addition of one known element (i.e., Iwase’s word replacement feature) to another known element (i.e., Shimomura’s distance-based output adjustment) to obtain predictable results (i.e., adjusting the output using word replacement). *Id.* (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶ 210). Amazon asserts that there would have been a reasonable expectation of success in incorporating Iwase’s word replacement feature into Shimomura’s control unit 10 because (1) “this configuration would have been trivial, involving straightforward changes to enhance the control unit’s approach to adjusting its output content”; and (2) “such an approach was well known.” *Id.* at 62 (citing Ex. 1002 ¶ 211; Ex. 1011).

In its Patent Owner Response, SoundClear relies on the same arguments and evidence it presents for the obviousness ground based on the combined teachings of Shin and Iwase to rebut Amazon’s explanation as to how the combined teachings of Shimomura and Iwase account for limitation 1[f]. *Compare* PO Resp. 53–55, *with id.* at 39–42. For the same reasons we identify above, we do not agree with SoundClear’s arguments. *See supra*

Section II.C.2. Accordingly, Amazon has demonstrated by a preponderance of the evidence that the subject matter of independent claims 1, 6, and 7 would have been obvious over the combined teachings of Shimomura and Iwase.

2. Claims 2 and 5

Beginning with dependent claim 2, SoundClear does not address separately Amazon's arguments and supporting evidence explaining how the combined teachings of Shimomura and Iwase account for the additional limitations of this dependent claim. *See* PO Resp. 53–55. We have reviewed Amazon's arguments and supporting evidence explaining how the combined teachings of Shimomura and Iwase account for the additional limitations of dependent claim 2, and we agree with and adopt Amazon's analysis. *See* Pet. 62. Accordingly, SoundClear has demonstrated by a preponderance of the evidence that the subject matter of dependent claim 2 would have been obvious over the combined teachings of Shimomura and Iwase.

Turning to dependent claim 5, Amazon relies on the same arguments and evidence it presents for the obviousness ground based on the combined teachings of Shimomura and Aoyama to support its obviousness ground based on the combined teachings of Shimomura and Iwase challenging this dependent claim. *See* Pet. 62. As we explain above, we agree with SoundClear that Shimomura does not teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See supra* Section II.D.4. In the Petition, Amazon does not rely on Iwase to cure this deficiency in Shimomura. *See* Pet. 62. Accordingly, SoundClear has not demonstrated by a preponderance of the evidence that the subject

matter of dependent claim 5 would have been obvious over the combined teachings of Shimomura and Iwase.

F. Obviousness over the Shin, Iwase, and Aoyama/Obviousness over Shimomura, Iwase, and Aoyama

Amazon contends that claims 3 and 4 of the '675 patent are unpatentable under 35 U.S.C. § 103 as obvious over 3(a) the combined teachings of Shin, Iwase, and Aoyama; and 3(b) the combined teachings of Shimomura, Iwase, and Aoyama.¹⁰ Pet. 63–65. Amazon contends that each of the proposed obviousness combinations teaches or suggests the subject matter of dependent claims 3 and 4, 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 for independent claim 1 to rebut the obviousness grounds based on the combined teachings of 1(a) Shin and Aoyama, 1(b) Shin and Iwase, 2(a) Shimomura and Aoyama, and 2(b) Shimomura and Iwase, SoundClear asserts that neither the combined teachings of Shin, Iwase, and Aoyama, nor the combined teachings of Shimomura, Iwase, and Aoyama, teach or suggest the additional limitations of dependent claims 3 and 4. PO Resp. 56. To support its positions,

¹⁰ For clarity and ease of reference, we rely on the numbering of asserted grounds of unpatentability provided by Amazon in the Petition. *See* Pet. 5 (Ground 3 is labeled as “Grounds 1b or 2b with Aoyama,” where Ground 1b is labeled as “Shin and Iwase” and Ground 2b is labeled as “Shimomura and Iwase”).

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 3 and 4 of the '675 patent are unpatentable as obvious over 3(a) the combined teachings of Shin, Iwase, and Aoyama; and 3(b) the combined teachings of Shimomura, Iwase, Aoyama.

Beginning with dependent claim 3, Amazon contends that Aoyama teaches or suggests the additional limitations required by this dependent claim. Pet. 63 (citing Ex. 1002 ¶¶ 217, 218). According to Amazon, it would have been obvious to configure Shin's processor 120 and/or Shimomura's control unit 10 to implement Iwase's word replacement feature using Aoyama's tables. *Id.* (citing Ex. 1002 ¶¶ 219–222). Amazon argues that there are several reasons as to why a person of ordinary skill in the art would have been motivated to configure Shin's processor 120 and/or Shimomura's control unit 10 “to use separate tables for storing expression variants corresponding to common semantic concepts disclosed in Iwase, such as ‘request,’ ‘proposal,’ ‘wish,’ or ‘opinion.’” *Id.* (citing Ex. 1002 ¶ 219). As one example, Amazon argues that configuring Shin's processor 120 and/or Shimomura's control unit 10 in this way reflects a combination of known elements (i.e., Shin's and Shimomura's distance-based output adjustments, Iwase's word replacement feature, and Aoyama's organization of replacement words in tables) to achieve predictable results (i.e., generating responses by looking up words in semantic concepts tables and replacing those words with more general ones from the tables). *Id.* at 63–64 (citing *KSR*, 555 U.S. at 417; Ex. 1002 ¶ 220). Amazon asserts that there

would have been a reasonable expectation of success in configuring Shin’s processor 120 and/or Shimomura’s control unit 10 to implement Iwase’s word replacement feature using Aoyama’s tables because “[t]his configuration would have been trivial, involving straightforward changes to enhance the processor’s [and control unit’s] approach to simplifying and abbreviating its output content.” *Id.* at 64 (citing Ex. 1002 ¶ 221).

Turning to dependent claim 4, Amazon contends that a person of ordinary skill in the art would have been motivated to combine the teachings of either Shin or Shimomura with those of Iwase and Aoyama because it would result in “us[ing] separate tables for storing expression variants corresponding to common semantic concepts.” Pet. 64 (citing Ex. 1002 ¶ 223). According to Amazon, this configuration would result in “each table of expression variant [being] associated with a distinct semantic concept like ‘request’ or ‘proposal,’” such that “each table includes type information (e.g., ‘request’ or ‘proposal’).” *Id.* at 64–65. Amazon further argues that, consistent with the disclosures in Iwase and Aoyama, “when a word in the notification information has a semantic concept that matches the semantic concept of a table and also appears in the table, the processor [or control unit] would replace the word in the notification with an expression variant from the table.” *Id.* at 65.

In its Patent Owner Response, apart from the arguments it presents for independent claim 1 to rebut the obviousness grounds based on the combined teachings of 1(a) Shin and Aoyama, 1(b) Shin and Iwase, 2(a) Shimomura and Aoyama, and 2(b) Shimomura and Iwase, SoundClear does not address separately Amazon’s arguments and supporting evidence explaining how the combined teachings of Shin, Iwase, and Aoyama, and

the combined teachings of Shimomura, Iwase, and Aoyama, account for the additional limitations of dependent claims 3 and 4. PO Resp. 56. For the same reasons we identify above, we do not agree with SoundClear's arguments. *See supra* Sections II.B.5.a–d, II.C.2, II.D.2.a–d, II.E.1. Moreover, we have reviewed Amazon's arguments and supporting evidence explaining how the combined teachings of Shin, Iwase, and Aoyama, and the combined teachings of Shimomura, Iwase, and Aoyama, account for the additional limitations of dependent claims 3 and 4, and we agree with and adopt Amazon's analysis. *See* Pet. 63–65. Accordingly, Amazon has demonstrated by a preponderance of the evidence that the subject matter of claims 3 and 4 would have been obvious over 3(a) the combined teachings of Shin, Iwase, and Aoyama; and 3(b) the combined teachings of Shimomura, Iwase, and Aoyama.

G. Obviousness Grounds Based, in Part, on Schuster

Amazon contends that claim 5 of the '675 patent is unpatentable under 35 U.S.C. § 103 as obvious over 4(a) the combined teachings of Shin, Aoyama, and Schuster; 4(b) the combined teachings of Shin, Iwase, and Schuster; 4(c) the combined teachings of Shimomura, Aoyama, and Schuster; and 4(d) the combined teachings of Shimomura, Iwase, and Schuster. Pet. 65–67. Amazon contends that each of the proposed obviousness combinations teaches or suggests the subject matter of dependent claim 5, 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.

SoundClear contends that Schuster does not teach or suggest whether the voice spoken by the user is a whisper, as required by dependent claim 5. PO Resp. 56–59. 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 claim 5 of the '675 patent is unpatentable as obvious over 4(a) the combined teachings of Shin, Aoyama, and Schuster; 4(b) the combined teachings of Shin, Iwase, and Schuster; 4(c) the combined teachings of Shimomura, Aoyama, and Schuster; and 4(d) the combined teachings of Shimomura, Iwase, and Schuster. We begin our analysis by providing an overview of Schuster, and then we turn to the parties' contentions with respect to this challenged claim.

1. Schuster

Schuster generally relates to “mobile communication devices” and, in particular, to “a system and method for modifying a communication mode of a mobile communication device based on user speech characteristics.” Ex. 1016, 1:7–11. Schuster discloses “a mobile communication device [that] enters a private communication mode with the user when the user’s preceding voice interaction with the device is whispered.” *Id.* at 2:9–62; *see also id.* at 5:37–41 (disclosing that “the device will move to the private mode state 203 in an embodiment when the device detects the intended private interaction from the user,” such as a “whisper detection”). Upon detecting a private interaction with the user (e.g., detecting a whisper) and entering a private communication mode, “this mode alters the way in which the device interacts with the user.” *Id.* at 3:31–33. For example, “the device

may significantly lower the volume of speech feedback so that only the user or someone very close to the device (e.g., within one or two feet) will be able to hear the response.” *Id.* at 3:35–39. According to Schuster, the private communication mode “allows the user to interact with the device in a private manner, avoid[s] disturbing other people, keep[s] [the] interaction confidential, or simply . . . avoid[s] [the] social awkwardness associated with communicating with an electronic device.” *Id.* at 2:63–67.

2. Claim 5

Claim 5 depends from independent claim 1 and further recites “the voice classifying unit is further configured to determine the voice to be the predetermined voice when the voice spoken by the user is a whisper.” Ex. 1001, 22:3–7.¹¹ Amazon contends that Schuster teaches or suggests this limitation because it discloses using microphone signal analysis to detect a whisper from a user and, upon detecting a whisper, enters a private communication mode that alters a device’s communication behavior with the user. Pet. 65–66 (citing Ex. 1016, Abstract, 1:7–11, 2:6–10, 2:59–67, 3:2–4, 3:12–18, 3:31–39, 6:34–46, 7:18–20, claims 3, 4, 12, 13, 19). Amazon then argues that, even if SoundClear were to take the position that dependent claim 5 requires determining whether the voice spoken by a user is a whisper, Schuster still teaches or suggest this limitation because it distinguishes “between a first voice and a second voice based on whether the voice is whispered or non-whispered.” *Id.* at 66 (citing Ex. 1002 ¶ 227).

Turning to rationale to combine and reasonable expectation of success, Amazon contends that there are several reasons why a person of

¹¹ See *supra* n.5.

ordinary skill in the art would have been motivated to configure Shin's processor 120 and/or Shimomura's control unit 10 to include Schuster's whisper detection. Pet. 66 (citing Ex. 1002 ¶¶ 228–231). As one example, Amazon argues that configuring Shin's processor 120 and/or Shimomura's control unit 10 in this way reflects nothing more than applying a known technique (i.e., Schuster's whisper detection) to improve a similar device (Shin's processor 120 and/or Shimomura's control unit 10) in the same way. *Id.* (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶¶ 229). Amazon asserts that there would have been a reasonable expectation of success in modifying Shin's processor 120 and/or Shimomura's control unit 10 to include Schuster's whisper detection because (1) "Shin[']s and Shimomura's devices already classify speech into two or more categories"; and (2) "[e]xtending their capabilities to process whispers and/or classify users based on whispered or non-whispered input would involve straightforward adjustments (if any) to the existing voice detection and processing hardware or software." *Id.* at 67 (citing Ex. 1002 ¶ 230).

In its Patent Owner Response, SoundClear contends that Amazon's proposed obviousness combinations based, in part, on Schuster are "insufficient" because Schuster "discloses an unrelated process" of "shifting between modes (e.g., a 'normal mode' and 'private mode')." PO Resp. 57 (quoting Ex. 1016, 5:37–39) (citing Ex. 2016 ¶ 157). SoundClear argues that Schuster does not determine a whisper in the context of classifying voices, and Schuster's device does not respond to the whisper detection—it merely responds based on mode. *Id.* (citing Ex. 1016, 3:31–33; Ex. 2016 ¶ 157). SoundClear further argues that "Schuster offers no guidance whatsoever on how a whisper could be detected based on distance at all—or

why a [person of ordinary skill in the art] would have been motivated to do so.” *Id.* at 58 (citing Ex. 2016 ¶ 158). Instead, SoundClear asserts that introducing distance to Schuster’s whisper detection “would risk confounding the determination” because a person of ordinary skill in the art would have recognized that “Schuster’s microphone signal analysis involves entirely different signals, metrics, and calculations than those in Shin[’s] and Shimomura[’s] proximity sensor calculation.” *Id.* (citing Ex. 2016 ¶ 158). According to SoundClear, a person of ordinary skill in the art “would not have known how to apply, and would have had no expectation of success in applying, Schuster’s teaching of detecting a whisper using a microphone to Shin[’s] and Shimomura’s distance-based output schemes.” *Id.*

In its Reply, Amazon contends that “[SoundClear] admits [that] Schuster discloses detecting [a] whisper” because, in response to a whisper detection, it changes modes and outputs different results. Pet. Reply 24 (citing PO Resp. 56–59; Ex. 1016, 3:31–39). Amazon further argues that it explained in the Petition why a person of ordinary skill in the art would have been motivated to implement Schuster’s whisper detection in Shin and/or Shimomura, and its declarant, Dr. Stern, testifies that doing so would have involved only “straightforward adjustments (if any) to the existing voice detection and processing hardware or software.” *Id.* at 24–25 (citing Ex. 1002 ¶¶ 229, 230). Amazon asserts that Schuster, itself, supports this proposed modification to Shin and/or Shimomura because it “classif[ies] based on both distance and whisper detection.” *Id.* at 25 (citing Ex. 1016 3:27–30).

In its Sur-reply, SoundClear maintains its argument that a person of ordinary skill in the art would not have been motivated by Schuster to

modify Shin and/or Shimomura to classify a voice based on whether it is a whisper because “Schuster discloses determining [a] whisper using a microphone for a process unrelated to classifying voices based on distance.” PO Sur-reply 20 (citing PO Resp. 56–59). According to SoundClear, Amazon does not explain in its Reply why a person of ordinary skill in the would have been motivated to modify Shin and/or Shimomura “based on Shuster’s disclosure of a process unrelated to modifying content output based on distance (like that described in Shin and Shimomura).” *Id.*

Based on the complete record, we agree with Amazon that Schuster teaches or suggests determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See* PO Resp. 65–67; Pet. Reply 24–25. As we explain in our claim construction analysis above, both the plain language of dependent claim 5 and the specification support requiring the voice classifying unit to determine whether the voice spoken by a user is a whisper. *See supra* II.A.2. The specification of the ’675 patent sets forth an explicit definition for the claim term “whisper” as “an unvoiced voice spoken without vibrating the vocal chords,” but it also includes a voice with a peak frequency equal to or lower than a threshold. *Id.* With this in mind, we address Amazon’s reliance on Schuster to teach or suggest determining whether the voice spoken by the user is a whisper, as required by dependent claim 5.

Schuster discloses “a mobile communication device [that] enters a private communication mode with the user when the user’s preceding voice interaction with the device is whispered.” Ex. 1016, 2:59–62; *see also id.* at 5:37–41 (disclosing that “the device will move to the private mode state 203 in an embodiment when the device detects the intended private interaction

from the user,” such as a “whisper detection”). According to Schuster, the device detects the intended private interaction with the user by “us[ing] one or a combination of several methods.” *Id.* at 3:1–2. One such method includes receiving a signal at a microphone and then “analyz[ing] it to detect a whispered microphone input.” *Id.* at 3:2–4. Schuster, however, makes clear that “any suitable method may be used . . . to detect whispered input.” *Id.* at 3:9–10. Schuster discloses that, upon detecting an intended private interaction with the user (i.e., by detecting a whisper), the device changes the mode “in which the device interacts with the user” (*id.* at 3:31–33) by entering a “private mode” (*id.* at 2:59–62, 5:37–41).

Based on these cited disclosures, Amazon argues—and we agree—that Schuster teaches or suggests receiving a voice spoken by a user at a microphone and analyzing the corresponding voice signal to determine whether it is a whisper. Pet. 65–66. This comports with our construction of the claim term “whisper” because Schuster’s whisper detection includes determining whether the voice spoken by the user is unvoiced speech, substantially unvoiced speech, or hushed speech in essentially the same way to the ’675 patent. Ex. 1016, 3:4–8 (“As used herein, the terms ‘whisper’ and ‘whispered’ refer to unvoiced or substantially unvoiced speech as well as hushed speech, *which may be spectrally distinct from either voiced speech or whispered speech.*” (emphasis added)).

We do not agree with SoundClear’s argument that Schuster detects a whisper using a microphone for a process unrelated to voice classification based on distance. *See* PO Resp. 57; PO Sur-reply 20. SoundClear’s argument improperly attacks Schuster in isolation. *See In re Keller*, 642 F.2d at 426; *In re Young*, 927 F.2d at 591. As we explain above, Amazon’s

proposed obviousness grounds rely on Shin and/or Shimomura—not Schuster—to teach or suggest the voice classification required by limitation 1[b][i]. *See* Pet. 12–15, 49–50. Amazon only relies on Schuster’s whisper detection to teach determining whether the voice spoken by the user is a whisper, as required by dependent claim 5. *See id.* at 65–67.

We also do not agree with SoundClear’s argument that a person of ordinary skill in the would not have been motivated to modify Shin’s processor 120 and/or Shimomura’s control unit 10 to include Schuster’s whisper detection. *See* PO Resp 58; PO Sur-reply 20. SoundClear’s argument in this regard is predicated on the notion that Schuster’s whisper detection is unrelated to Shin’s and Shimomura’s content output schemes that are based on user-to-device distance. This is incorrect. Schuster discloses that, “additionally, other sensors may be used as well to detect the user’s desire to enter the private mode, for example, proximity sensors can be used to sense proximity of the user.” Ex. 1016, 3:27–30. Therefore, similar to Shin and Shimomura, Schuster also contemplates using proximity sensors to measure user-to-device distance to determine how its device interacts with a user. In our view, this serves to bolsters Amazon’s stated rationales to modify Shin’s processor 120 and/or Shimomura’s control unit 10 to include Schuster’s whisper detection. *See* Pet. 66–67; Ex. 1002 ¶¶ 228–231.

In summary, Amazon has demonstrated by a preponderance of the evidence that the subject matter of dependent claim 5 would have been obvious over 4(a) the combined teachings of Shin, Aoyama, and Schuster; 4(b) the combined teachings of Shin, Iwase, and Schuster; 4(c) the combined

teachings of Shimomura, Aoyama, and Schuster; and 4(d) the combined teachings of Shimomura, Iwase, and Schuster.

H. Obviousness Grounds Based, in Part, on Kristjansson

Amazon contends that claims 1–7 of the '675 patent are unpatentable under 35 U.S.C. § 103 as obvious over 5(a) the combined teachings of Shin, Aoyama, and Kristjansson; 5(b) the combined teachings of Shin, Iwase, and Kristjansson; 5(c) the combined teachings of Shimomura, Aoyama, and Kristjansson; 5(d) the combined teachings of Shimomura, Iwase, and Kristjansson; 5(e) the combined teachings of Shin, Iwase, Aoyama, and Kristjansson; 5(f) the combined teaching of Shimomura, Iwase, Aoyama, and Kristjansson; 5(g) the combined teachings of Shin, Aoyama, Schuster and Kristjansson; 5(h) the combined teachings of Shin, Iwase, Schuster, and Kristjansson; 5(i) the combined teachings of Shimomura, Aoyama, Schuster, and Kristjansson; and 5(j) the combined teachings of Shimomura, Iwase, Schuster, and Kristjansson. Pet. 67–69.

Section 312(a)(3) of Title 35 of the United States Code requires a petition 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). Amazon’s application of Kristjansson to all the obviousness grounds raised in the Petition would result in at least ten different combinations of asserted prior art references. Amazon, however, does not present a mapping of the teachings of each prior art reference that Amazon proposes to combine to account for each and every limitation of claims 1–7. For this reason alone, we determine that the Petition fails to meet the particularity requirement of 35 U.S.C. § 312(a)(3) for all the obviousness grounds based, in part, on Kristjansson.

Accordingly, Amazon has not demonstrated by a preponderance of the evidence that the subject matter of claims 1–7 would have been obvious over 5(a) the combined teachings of Shin, Aoyama, and Kristjansson; 5(b) the combined teachings of Shin, Iwase, and Kristjansson; 5(c) the combined teachings of Shimomura, Aoyama, and Kristjansson; 5(d) the combined teachings of Shimomura, Iwase, and Kristjansson; 5(e) the combined teachings of Shin, Iwase, Aoyama, and Kristjansson; 5(f) the combined teaching of Shimomura, Iwase, Aoyama, and Kristjansson; 5(g) the combined teachings of Shin, Aoyama, Schuster and Kristjansson; 5(h) the combined teachings of Shin, Iwase, Schuster, and Kristjansson; 5(i) the combined teachings of Shimomura, Aoyama, Schuster, and Kristjansson; and 5(j) the combined teachings of Shimomura, Iwase, Schuster, and Kristjansson.

I. 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. 12–23, 60–62. We address SoundClear’s arguments in turn.

1. Equating Measurements to the Object Being Measured

SoundClear contends that all the obviousness grounds raised in the Petition fail because they improperly equate the concept of classifying an object with “manners of measuring variables, like distance.” PO Resp. 12. Relying on the deposition testimony of Amazon’s declarant, Dr. Stern, SoundClear asserts that Shin and/or Shimura, at best, classify distances and not voices. *Id.* at 12–13 (citing Ex. 2017, 48:3–7, 49:7–15, 95:7–17, 117:1–21, 118:2–13, 119:6–22). SoundClear further argues that the ’675 patent

treats the concept of “volume” as distinct from “voice,” and “[n]one of the prior art discloses classifying a voice separate from measuring a characteristic of the voice.” *Id.* at 13 (citing Ex. 1001, 2:33–36, 10:9–46). SoundClear also argues that Shin and/or Shimomura each identify measurements and processing as completely independent of one another, and the Petition “never bridges the gap in the context of classifying an input separate and apart from adjusting an output.” *Id.* at 14 (citing Ex. 1003, Table 3 as distinct from Tables 1, 2, or 4–9; Ex. 1004, Fig. 7 as distinct from Figs. 9–12). Lastly, SoundClear asserts that “Shin and Shimomura disclose specific analyses of voice in simplistic fashion,” each of which “relies on a single-variable table,” but Amazon does not explain why a person of ordinary skill in the art would modify Shin and/or Shimomura to include “a process of voice classification.” *Id.* at 15 (citing Pet. 13, 50; Ex. 2016 ¶ 76).

In its Reply, Amazon maintains that Shin and Shimomura each disclose voice classification based on distance. Pet. Reply 4. Amazon contends that the deposition testimony of Dr. Stern supports its position in this regard because he “repeatedly explained that the prior art discloses *classifying voice based on distance*.” *Id.* at 4–5 (citing Ex. 2017, 57:6–13, 57:18–58:5, 60:13–61:4). Amazon also disagrees with SoundClear’s argument that Shin and/or Shimomura only classify distances and not voice because “[they] disclose[] devices that acquire a user’s voice, calculate the use-to-device distance, and classify the voice based on the distance—exactly as the ’675 patent does.” *Id.* at 5 (citing Ex. 1202 ¶ 8). Amazon further argues that the prior art does not treat measurements and processing as completely independent of one another, but rather discloses that they are

used together. *Id.* For example, Amazon asserts that Shin explicitly discloses using Tables 1 and 3 together, and that Shimomura’s process of using distance-based classification to determine whether to alter the content output to the user serves as the input to using distance-based classification to determine whether to alter the output volume. *Id.* at 5–6 (citing Ex. 1003 ¶¶ 49, 53, 140, 183, 184, claim 17; Ex. 1004 ¶¶ 56–70, Figs. 7, 8, 9, 13; Ex. 1202 ¶ 9).

In its Sur-reply, SoundClear contends that, even if one were to equate “classification” and “measurement,” “a ‘voice classification’ cannot be based solely on ‘distance.’” PO Sur-Reply 2. Relying on the deposition testimony of Amazon’s declarant, Dr. Stern, from a related IPR, SoundClear maintains that “voice” is distinct from “distance.” *Id.* at 2–3 (citing Ex. 2020, 28:12–18, 31:11–32:6, 54:1–6). SoundClear argues that, “unlike ‘feature values’ that directly bear on voice attributes/characteristics (e.g., peak frequency), distance has no direct relationship with any voice attributes/characteristics.” *Id.* at 3. SoundClear asserts that, “[t]o the extent a distance threshold must be set in a manner fit for this purpose—the threshold cannot be consistent with more than one class of voices.” *Id.*

In essence, SoundClear’s arguments that the obviousness grounds raised in the Petition fail because they improperly equate the concept of classifying a voice with measuring variables, like distance, effectively rehashes the same arguments SoundClear raised to rebut the obviousness grounds based, in part, on Shin, and the obviousness ground based on, in part, on Shimomura. For the same reasons we identify above, we do agree with SoundClear’s arguments. *See supra* Section II.B.5.a; Section II.D.2.a.

At best, we understand SoundClear to argue that the challenged claims require distance to be used in conjunction with other feature values to classify a voice. *See* PO Resp. 13; PO Sur-reply 3. But as we explain in our claim construction analysis above, the plain language of the claim phrase “classify the voice into either a first voice or a second voice based on the calculated distance” controls because it merely requires classifying a voice spoken by a user that is acquired by a device as either a first or second voice based on the distance between the user and the device. *See supra* Section II.A.1. Although the specification of the ’675 patent contemplates using feature values other than distance to perform the claimed voice classification, the plain language of the challenged claims only requires performing the claimed voice classification using one feature value—specifically, distance. *See* Ex. 1001, 9:58–10:46.

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, Shin still teaches or suggests this unduly narrow reading of limitation 1[b][ii]. That is, Shin’s processor 120 would still effectively perform the claimed voice classification 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.

2. Failure to Address Obviousness Under Prevailing Precedent

In the Patent Owner Response, SoundClear contends that the two rationales to combine highlighted in our Institution Decision relied on the same type of conclusory statements that the U.S. Court of Appeals for the

Federal Circuit determined to be insufficient in other cases. PO Resp. 16–17 (citing Dec. on Inst. 26–27). According to SoundClear, neither of these rationales highlighted in our Institution Decision “address the full context of the claims because neither addresses classification of voices separate from changing outputs.” *Id.* at 17 (citing Ex. 2016 ¶ 78, 79). In its Reply, Amazon contends that the rationales to combine identified by SoundClear in our Institution Decision do not simply identify elements known in the prior art, but rather they explain “what elements would be combined . . . and what beneficial, predictable results would be produced.” Pet. Reply 6–7 (citing Dec. on Inst. 26–27; Pet. 19–20). In addition, Amazon argues that SoundClear “ignores the many additional motivations the Petition provided.” *Id.* at 7 (citing Pet. 19–20, 25–27).

We do not agree with SoundClear’s argument that Amazon’s rationales to combine are conclusory and do not address the full context of the challenged claims. In our view, SoundClear merely provides snippets of Amazon’s rationale to combine arguments in isolation, which we reiterated in the Decision on Institute, but fails to appreciate the reference-specific nature of those arguments. *See* PO Resp. 16–17. This approach does not convince us that Amazon’s rationales to combine set forth in the Petition represent the type of conclusory statements that Federal Circuit has deemed insufficient to support a conclusion of obviousness.

In any event, SoundClear only addresses two of seven rationales to combine set forth in the Petition that Amazon provides to justify configuring Shin’s processor 120 to adjust output information based on the user’s distance from electronic device 101 by incorporating Aoyama’s word replacement feature. *See* Pet. 25–28. More importantly, as we explain

above, Amazon argues—and we agree—that configuring Shin’s processor 120 in this way compliments Shin’s goal of abbreviating or simplifying the content output to a user at a greater distance from electronic device 101 to make the content easier for the user to understand. *See, e.g.*, Ex. 1003 ¶ 89 (disclosing that “a user who requests content with a voice input to the electronic device 101 while being far away from the electronic device 101 may want core content corresponding to the voice input rather than detailed content”).

3. *Reliance on Hindsight*

In its Patent Owner Response, SoundClear directs us to the deposition testimony of Amazon’s declarant, Dr. Stern, to support its argument that Amazon relies on impermissible hindsight reconstruction. PO Resp. 19 (citing Ex. 2017, 59:9–60:5, 67:7–15). According to SoundClear, this “alone undermines the opinions in [Dr. Stern’s] declarations and the grounds in the Petition that rely on those opinions.” *Id.* SoundClear asserts that Amazon does not “meet [its] burden to provide some reason for the proposed combinations,” which “is the only way to avoid clear hindsight bias that [Amazon] and their [declarant] used to build the Petition.” *Id.* at 20. In its Reply, Amazon contends that Dr. Stern never admitted to relying on impermissible hindsight reconstruction. Pet. Reply 7 (citing PO Resp. 19). Instead, Amazon asserts that Dr. Stern explained that “with 20/20 hindsight” “he might have *provided slightly different annotations in his declaration* for greater clarity.” *Id.* (quoting Ex. 2017, 59:22–60:5, 67:11–15).

We do not agree with SoundClear’s argument that Amazon relies on impermissible hindsight reconstruction. SoundClear mischaracterizes

Dr. Stern's deposition testimony to support its hindsight argument.

Dr. Stern never admitted to relying on impermissible hindsight reconstruction when providing testimony in support of Amazon's proposed obviousness grounds. As one example, when counsel for SoundClear inquired as to why paragraph 61 of Dr. Stern's Declaration accompanying the Petition included a citation to paragraph 51 of Shin, together with a corresponding parenthetical that included content may be "classified dichotomously," Dr. Stern responded by indicating "with 20/20 hindsight, in annotating Table 1 [of Shin], it probably would have been better to have just led with the first voice being any distance greater than 1 meter as, in fact, I do suggest at the tail end of paragraph 61." Ex. 2017, 59:4–60:6. We agree with Amazon that Dr. Stern's reference to "20/20 hindsight" indicated his desire to provide slightly different annotations in his Declaration accompanying the Petition for purposes of greater clarity. *See* Pet. Reply 7. Dr. Stern's reference to "20/20 hindsight" does not amount to an admission that he engaged in impermissible hindsight reconstruction.

4. Dr. Stern's Testimony is Reliable

In its Patent Owner Response, SoundClear contends that "Dr. Stern's [D]eclaration is unreliable" because "[h]is positions are . . . malleable, and he defaults to inferring certain disclosures in the prior art without any basis." PO Resp. 21. SoundClear then directs us to several examples of Dr. Stern's declaration and deposition testimony that SoundClear asserts should be given little to no weight. *Id.* at 21–23 (citing Ex. 2017, 210:4–10, 6:21–27:5, 48:3–7, 49:7–15, 55:8–57:5, 57:18–60:12, 62:3–63:18, 65:11–22, 72:6–75:14, 96:8–9, 102:9–10, 103:1–16, 119:6–22, 120:3–13, 117:1–21; Ex. 1002 ¶ 61). In its Reply, Amazon contends that Dr. Stern "properly

analyze[s] the claims and prior art.” Pet. Reply 8. Amazon then proceeds to walk through each of the examples of Dr. Stern’s declaration and deposition testimony identified by Amazon in the Patent Owner Response and explain why “each allegation is belied by [his] testimony.” *Id.* at 8–10.

We do not agree with SoundClear that Dr. Stern’s testimony is unreliable. Even if SoundClear were correct as to deficiencies in certain portions of Dr. Stern’s declaration and deposition testimony, that would not necessarily result in us discounting his testimony in its entirety. *See, e.g., Velandier v. Garner*, 348 F.3d 1359, 1371 (Fed. Cir. 2003) (“It is within the discretion of the trier of fact to give each item of evidence such weight as it feels appropriate.”); *Yorkey v. Diab*, 601 F.3d 1279, 1284 (Fed. Cir. 2010) (“[T]he Board was well within its discretion to give more credibility to [one portion of expert testimony over another] unless no reasonable trier of fact could have done so.”). 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.”).

5. *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. 60–61 (citing Ex. 2016 ¶¶ 164, 165). 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.* at 61. In its Reply,

Amazon disagrees with SoundClear’s argument that Shin and Shimomura both describe simpler systems than the ’675 patent. Pet. Reply 25.

According to Amazon, Shin and Shimomura each “disclose the very system the ’675 patent describes and claims.” *Id.* (citing Ex. 1202 ¶ 10).

As an initial matter, SoundClear repeats essentially the same arguments it presents to rebut the obviousness grounds based, in part, on Shin, and the obviousness grounds based, in part, on Shimura. For the same reasons we identify above, we do not agree with SoundClear’s argument. *See supra* Section II.B.5.a; Section II.D.2.a. In addition, we agree with Amazon that Shin and Shimomura each teach or suggest classifying voice input by using distance in the same way as the ’675 patent (i.e., by using distance thresholds between a user and a device). *Compare* Ex. 1003 ¶¶ 86, 87, Table 3, Ex. 1004 ¶¶ 60–64, Fig. 7, *with* Ex. 1001, 9:60–10:46.

6. *Graham Factor Two*

SoundClear contends that “[Amazon] fail[s] to meet *Graham* factor two in its obviousness analysis” because Amazon does “not explain what specifically is missing from the primary references, where it is found, or why a [person of ordinary skill in the art] would have been motivated to bridge the gap.” PO Resp. 61. Instead, SoundClear asserts that Amazon fails to establish a *prima facie* case of obviousness because it only addresses *Graham* factor two in limited and isolated instances. *Id.* at 62. In its Reply, Amazon contends that SoundClear previously raised the same argument that the Petition fails to apply the *Graham* factors in two other papers filed during the preliminary phase in this proceeding and, in both instances, that argument was not found to be persuasive. Pet. Reply 25 (citing Paper 5, 44–49; Prelim. Resp. 27–31).

Beginning with Shin, the only differences between Shin and representative independent claim 1 applies to limitations 1[c], 1[d], and the word replacement required by limitation 1[f]. Amazon, however, specifically identifies these differences in the Petition and then explains how Aoyama teaches or suggests limitations 1[c], 1[d], and the word replacement required by limitation 1[f] or, alternatively, how Iwase teaches or suggests the word replacement required by limitation 1[f]. *See* Pet. 15–20, 22–28, 41–44. Turning to Shimomura, the only differences between the teachings of Shimomura and representative independent claim 1 applies to limitations 1[c], 1[d], and the word replacement required by limitation 1[f]. Amazon, however, specifically identifies these differences in the Petition and then explains how Aoyama teaches or suggest limitations 1[c], 1[d], and the word replacement required by limitation 1[f] missing from Shimomura or, alternatively, how Iwase teaches or suggests the word replacement required by limitation 1[f] missing from Shimomura. *See* Pet. 51–57, 60–62. Bottom line, we are convinced that Amazon correctly applies *Graham* factor two in its obviousness analysis of Shin and Shimura, which, on their own, serve as the base prior art reference for all the obviousness grounds raised in the Petition.

III. CONCLUSIONS¹²

Based on the fully developed trial record, Amazon demonstrates by a preponderance of the evidence that claims 1–7 of the '675 patent are unpatentable. A summary of our conclusions is set forth in the tables below.

¹² 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
1–7	103	Shin, Aoyama	1–5, 6, 7	5
1, 2, 5–7	103	Shin, Iwase	1, 2, 6, 7	5
1–7	103	Shimomura, Aoyama	1–4, 6, 7	5
1, 2, 5–7	103	Shimomura, Iwase	1, 2, 6, 7	5
3, 4	103	Shin, Iwase, Aoyama	3, 4	
3, 4	103	Shimomura, Iwase, Aoyama	3, 4	
5		Shin, Aoyama, Schuster	5	
5		Shin, Iwase, Schuster	5	
5		Shimomura, Aoyama, Schuster	5	
5		Shimomura, Iwase, Schuster	5	
1–7		Shin, Aoyama, Kristjansson		1–7
1–7		Shin, Iwase, Kristjansson		1–7
1–7		Shimomura, Aoyama, Kristjansson		1–7
1–7		Shimomura, Iwase, Kristjansson		1–7
1–7		Shin, Iwase, Aoyama, Kristjansson		1–7

Claims	35 U.S.C. §	Reference(s)/ Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1–7		Shimomura, Iwase, Aoyama, Kristjansson		1–7
1–7		Shin, Aoyama, Schuster, Kristjansson		1–7
1–7		Shin, Iwase, Schuster, Kristjansson		1–7
1–7		Shimomura, Aoyama, Schuster, Kristjansson		1–7
1–7		Shimomura, Iwase, Schuster, Kristjansson		1–7
Overall Outcome			1–7	

IV. ORDER

Accordingly, it is

ORDERED that claims 1–7 of the '675 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.

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.

IPR2025-00673
Patent 11,244,675 B2

McKONE, *Administrative Patent Judge*, dissenting.

I dissent because I would not find that any of the asserted references or combinations teaches “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence *in which at least one word selected among words included in the content information of the notification information is replaced with another word*,” as recited in claim 1, and recited similarly in claims 6 and 7.¹³

¹³ I agree with the Majority’s resolutions of the remaining issues in the case.

Shin teaches setting the amount of content output to an “abbreviation level” based on the distance between the user and the electronic device; in particular, the device will output more detailed weather information at a close distance (e.g., including a summary, temperature, rainfall amount, and advice) and will omit all but the summary at a farther distance. Ex. 1003 ¶¶ 86–88. As Patent Owner observes (PO Resp. 29–30), Shin does not teach abbreviating content by replacing words. Petitioner argues that “[a person of ordinary skill in the art] would have readily understood Shin’s disclosure of abbreviating content to describe it more simply or provide the ‘gist’¹⁴] includes replacing words.” Pet. 23 (citing Ex. 1003 ¶¶ 51, 87; Ex. 1002 ¶ 92); *see also id.* at 27 (“Shin’s device presents only two design options for abbreviating or simplifying its output: omitting words or replacing them. Given Shin’s disclosure of simplifying output to provide ‘core content’ rather than ‘detailed content’ for users farther away, it would have been obvious to try a word-replacement approach.” (citing Ex. 1003 ¶ 89; Ex. 1002 ¶ 101)). Petitioner does not appear to contend that Shin expressly discloses word replacement (it does not); rather, Petitioner appears to argue that a skilled artisan would have inferred word replacement as an option or found it obvious. As Shin itself does not describe word replacement, Petitioner’s only evidence to support such an inference by a skilled artisan is the testimony of Dr. Stern, who repeats the Petition’s argument without

¹⁴ The “gist” refers to paragraph 51, describing “content that includes only a gist of the response corresponding to the voice input (a related example is described later herein).” As the “related example” clearly refers to omitting portions of output information, rather than replacing words, this “gist” description does not add anything to Petitioner’s evidence.

identifying the basis for the testimony. Ex. 1002 ¶¶ 92, 101. I would assign this “*verbatim* restatement of the assertion being supported, without any supporting evidence or technical reasoning” little weight and would find the argument unpersuasive. *See Xerox Corp. v. Bytemark, Inc.*, IPR2022-00624, Paper 9 at 16 (PTAB Aug. 24, 2022) (precedential). Thus, I would not find that Shin, by itself, teaches “at least one word . . . is replaced with another word,” as recited in claim 1.

As to Shimomura, it teaches inserting or removing words like “ne” after phrases based on distance between the user and the device. Pet. 54–55 (citing Ex. 1004 ¶¶ 57–66, 85, 92). Petitioner does not contend (and I would not find) that Shimomura teaches replacing words. *Id.* at 55 (citing only Aoyama for replacing words in output content).

Petitioner alleges that Shin and/or Shimomura, when combined with Aoyama, teaches word replacement based on distance (with Shin/Shimomura supplying adjusting the amount of information based on distance and Aoyama supplying word replacement as a means for information adjustment). Pet. 22–25, 54–55. I would not find that Aoyama remedies the deficiencies of either Shin or Shimomura because I would not find that Petitioner has articulated a rational reason to combine Aoyama with either reference.

Aoyama teaches rephrasing output content to comport with the “speech style” of the user. Ex. 1010 ¶¶ 138–140, 177, 218, 234–236. For example, Aoyama’s device might output “1 pm” instead of “13:00,” “10/1/2014” instead of “tomorrow,” or “Mr. Yamada” instead of “Taro Yamada,” depending on context and how the user typically would phrase this information. *Id.* ¶¶ 138–140. Petitioner argues that “Aoyama’s device

tailors output by replacing words based on the user or situation.” Pet. 23 (citing Ex. 1010 ¶¶ 120–124, 138–140, 218). To the extent rephrasing output to comport with speech style is word replacement, Aoyama does not describe (and Petitioner does not contend that Aoyama describes) replacing words based on the distance between the user and the device. Aoyama describes another example in which the system ascertains a “degree of noise” or “degree of composure” of the user, e.g., that the user is calm, hurried, running, walking, or lying down. Ex. 1010 ¶¶ 208–214. Here, however, Aoyama omits output content for lower degrees of composure, much like Shin describes omitting phrases when distance increases, and does not describe any behavior that could be equated to replacing words based on degree of composure. *Id.* ¶¶ 144–147; PO Resp. 31.¹⁵

As to Shin and Aoyama, Petitioner argues that “[a person of ordinary skill in the art] would have been motivated to configure Shin’s processor to adjust the information provided based on the user-to-device distance by incorporating Aoyama’s word replacement feature for many reasons.” Pet. 25 (citing Ex. 1002 ¶¶ 96–103). First, Petitioner argues that “replacing full names with simple nicknames” and “replacing complicated date and time formats with simpler terms” would “make the output easier to understand at greater distances.” *Id.* (citing Ex. 1002 ¶ 97). Aoyama does not teach that such replacement would have been easier to understand at greater distances. Rather, Aoyama teaches rephrasing information to correspond to speech style, which would be appropriate at any distance.

¹⁵ I do not agree with Patent Owner’s argument (PO Resp. 32) that Aoyama teaches away from replacing words based on distance.

Ex. 1010 ¶¶ 138–140. Instead, Petitioner relies only on Dr. Stern’s testimony to support this argument, and that testimony merely repeats the argument in the Petition, without identifying a basis. Ex. 1002 ¶ 97. I would not find this credible or persuasive. *See Xerox*, IPR2022-00624, Paper No. 9 at 16.

As to Aoyama’s degree of composure embodiment, Petitioner argues that a skilled artisan would have been motivated “to use Shin’s proximity sensor to replace words based on the user’s distance, which a [person of ordinary skill in the art] would have understood to be another example of user state” and would have made user-device interaction “more natural.” Pet. 25–26 (citing Ex. 1010 ¶¶ 97–114, 254; Ex. 1002 ¶ 98). Petitioner again relies only on expert testimony that repeats the Petition and does not identify a basis, which I would find lacking in credibility and unpersuasive. Ex. 1002 ¶ 98; *Xerox*, IPR2022-00624, Paper No. 9 at 16. Moreover, this embodiment of Aoyama is inapposite, as Aoyama does not teach replacing words to deal with lack of composure; rather, Aoyama teaches omitting information, similar to what Shin teaches. Ex. 1010 ¶¶ 144–147.

Next, Petitioner cites to Iyer (a reference not formally asserted by Petitioner as part of an obviousness combination) to argue that privacy concerns, such as the possibility of a passerby hearing the device’s confidential audible output, would have motivated a skilled artisan to replace words containing personal information with neutral alternatives. Pet. 26 (citing Ex. 1012, 3:8–15; Ex. 1002 ¶ 99). Iyer teaches transitioning between a normal mode and a discreet mode, for example diverting audible responses to an ear-piece speaker from a loudspeaker so that the user can send voice commands and receive responses at a lower volume. Ex. 1012, 3:8–15,

3:58–64. I would not find Petitioner’s argument based on Iyer persuasive, at least because neither Shin nor Aoyama teaches replacing personal information with neutral words—not for privacy, nor based on distance, nor for any other reason. Here, neither Shin nor Aoyama teaches a solution to the problem purportedly identified by Iyer.

KSR explains that “if a technique has been used to improve one device, and a person of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond his or her skill.” *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007). I am not persuaded by Petitioner’s evidence and argument that a skilled artisan would have recognized Aoyama to provide word replacement as a solution that simplifies/reduces the amount of information conveyed, either for privacy purposes or to make information easier to understand. Thus, in my view, a skilled artisan would not have recognized, based on Aoyama, that words output by Shin’s device could be replaced with other words to reduce information for understandability or to make information private when distance between the user and the device increases. In short, Aoyama does not teach word replacement providing the improvement Petitioner asserts; thus, the evidence does not show that Shin would have been improved “in the same way.” *KSR*, 550 U.S. at 417.

Petitioner’s remaining arguments merely paraphrase from *KSR*, without meaningful analysis or evidence (other than Dr. Stern’s testimony, which merely repeats the Petition without identifying a basis for the testimony). Pet. 26–27 (citing *KSR*, 550 U.S. at 417; Ex. 1002 ¶ 100). I do not find them persuasive.

Petitioner's reasons for combining Shimomura and Aoyama mimic those given for Shin and Aoyama, and are unpersuasive to me for the same reasons. Pet. 55–56 (citing Ex. 1002 ¶¶ 187–189). For example, Petitioner argues that a skilled artisan “would have understood that one application of Aoyama's word replacement feature would be to replace harder-to-understand formal names with easier-to-understand informal names, motivating a [person of ordinary skill in the art] to use the feature to advance Shimomura's goals.” *Id.* at 55 (citing Ex. 1004 ¶¶ 4–17; Ex. 1002 ¶ 187). Yet Aoyama does not describe replacing formal names with informal names for ease of understanding. Thus, Petitioner's only evidence is Dr. Stern's conclusory and unsupported testimony, which I do not find credible or persuasive. Ex. 1002 ¶ 187.

I also would not find that Iwase remedies the deficiencies of either Shin or Shimomura because I would not find that Petitioner has articulated a rational reason to combine Iwase with either reference.

Iwase describe extracting a subject on the basis of user attribute information and “shap[ing]” that subject for a target user to employ a speaking pattern that “is used by users in daily life,” for example, “learning the word order and words that are used by users for preference.” Ex. 1017, 15:6–14, 15:29–34. In one example, “where the subject is ‘I recommend bringing an umbrella.’, and a target user is a child, the speaking control unit 360 may shape the subject to be ‘You should take an umbrella.’ by using more direct expression.” *Id.* at 15:11–14. I share Patent Owner's concern that Iwase does not characterize its subject shaping as word replacement and Petitioner does not provide any meaningful explanation why it should be

considered replacement. PO Resp. 40.¹⁶ In any case, I would not find that Petitioner has shown a reason to combine Iwase with Shin or Shimomura.

Petitioner argues that

[a person of ordinary skill in the art] would have been motivated to configure Shin’s processor to adjust the information provided to the user based on the user-to-device distance by incorporating Iwase’s word replacement feature, for example to replace harder-to-understand words with simpler ones when the user is far away, for several reasons.

Pet. 42 (citing Ex. 1002 ¶¶ 145–151); *see also* Pet. Reply 20. I do not find any of these reasons persuasive.

Petitioner argues that Iwase teaches that its subject shaping allows a user to grasp notification contents more effectively, and that a skilled artisan would have incorporated Iwase’s word replacement feature into Shin’s device to gain this benefit. Pet. 42 (citing Ex. 1017, 2:11–14, 15:24–28, 21:66–22:8; Ex. 1002 ¶ 146); Pet. Reply 20. Petitioner elaborates on this in the Reply, arguing that “Petitioners and Dr. Stern explained that a [person of ordinary skill in the art] would have configured Shin’s processor to adjust the output based on the user-to-device distance by incorporating Iwase’s word replacement, for example, to replace harder-to-understand words with simpler ones when the user is far away.” Pet. Reply 20 (citing Ex. 1002 ¶¶ 145–151). This, however, is not what Iwase teaches. Instead, Iwase teaches changing the order in which information is presented in response to a high noise level, for example setting a subject output position to the latter

¹⁶ Petitioner’s response here is to incorrectly allege that “[Patent Owner] admits Iwase discloses ‘word replacement.’” Pet. Reply 19 (citing PO Resp. 39–40).

half of notification contents, rather than the earlier half, in order to get the user's full attention before the subject is presented. Ex. 1017, 4:30–41, 13:34–45. Iwase's other example, shaping the subject, is for the purpose of employing a pattern used in daily life, not to account for noise or distance. *Id.* at 15:4–40. Petitioner's remaining arguments merely paraphrase *KSR*, allege general similarities between Shin and Iwase, or repeat arguments presented for Shin, without providing meaningful explanation. Pet. 42–44. Dr. Stern merely repeats these conclusory arguments without explaining the basis for his testimony. Ex. 1002 ¶¶ 147–150. I would not find this testimony credible or these arguments persuasive.

Petitioner's arguments and evidence for combining Shimomura and Iwase are substantially the same as those presented for Shin and Iwase. Pet. 60–62 (citing Ex. 1002 ¶¶ 208–212). I would find them unpersuasive for the same reasons.

Since I would not find that any of Petitioner's grounds teaches “wherein the output-content generating unit is further configured to generate, when the voice is determined to be the first voice, a first output sentence *in which at least one word selected among words included in the content information of the notification information is replaced with another word,*” I respectfully dissent.

IPR2025-00673
Patent 11,244,675 B2
For PETITIONER

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

For PATENT OWNER

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