Trials@uspto.gov
571-272-7822

Paper 33
Entered: July 14, 2023

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SAMSUNG ELECTRONICS CO., LTD. and
SAMSUNG ELECTRONICS AMERICA, INC.,
Petitioner,

v.

STATON TECHIYA, LLC,
Patent Owner.

IPR2022-00253
Patent 9,491,542 B2

Before NATHAN A. ENGELS, SCOTT B. HOWARD, and RUSSELL E. CASS, *Administrative Patent Judges*.

CASS, *Administrative Patent Judge*.

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

I. INTRODUCTION

In this *inter partes* review, Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. ("Petitioner") challenge the patentability of claims 1–20 (the "challenged claims") of U.S. Patent No. 9,491,542 B2 (Ex. 1001, "the '542 patent"), which is assigned to Staton Techiya, LLC. ("Patent Owner").

We have jurisdiction under 35 U.S.C. § 6. This Final Written Decision, issued pursuant to 35 U.S.C. § 318(a), addresses issues and arguments raised during the trial in this *inter partes* review. For the reasons discussed below, Petitioner has proven by a preponderance of the evidence that claims 1–9 and 11–18 are unpatentable. Petitioner has not proven by a preponderance of the evidence that claims 10, 19, and 20 are unpatentable.

A. Procedural History

In this proceeding, Petitioner relies upon the following references:

> U.S. Patent Application Publication 2007/0189544 A1, published August 16, 2007 (Ex. 1005, "Rosenberg");
>
> U.S. Patent Application Publication 2011/0096939 A1, published April 28, 2011 (Ex. 1006, "Ichimura");
>
> U.S. Patent Application Publication 2011/0264447 A1, filed on April 22, 2011 and published October 27, 2011 (Ex. 1007, "Visser");
>
> U.S. Patent Application Publication 2010/0296668 A1, published November 25, 2010 (Ex. 1008, "Lee"); and
>
> U.S. Patent No. 6,728,385 B2, issued April 27, 2004 (Ex. 1009, "Kvaløy").

Pet. i–ii, 2. Petitioner also submits and relies upon declarations from of Dr. Nathanial Polish (Exs. 1002, 1028). Patent Owner submits and relies upon Declarations from David Kleinschmidt (Exs. 2001, 2006).

Petitioner challenges the patentability of claims 1–20 of the '542 patent based on the following grounds:

Ground	Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
1A	1–6	103	Rosenberg
1B	1–6	103	Rosenberg, Ichimura
2A	7–9	103	Rosenberg, Visser
2B	7–9	103	Rosenberg, Lee
3	10	103	Rosenberg, Visser, Kvaløy
4A	13–20	103	Rosenberg, Kvaløy, Visser
4B	13–20	103	Rosenberg, Kvaløy, Visser, Ichimura

Pet. i–ii, 3. Patent Owner filed a Preliminary Response. Paper 8. With our permission, Petitioner filed a Preliminary Reply to the Preliminary Response (Paper 9), and Patent Owner filed a Preliminary Sur-reply (Paper 10). We instituted trial on all grounds of unpatentability. Paper 13 ("Inst. Dec.") 25.

During the trial, Patent Owner filed a Response (Paper 19, "PO Resp."), Petitioner filed a Reply (Paper 21, "Pet. Reply"), and Patent Owner filed a Sur-reply (Paper 23, "PO Sur-reply"). With our permission, Petitioner filed a Sur-sur-reply (Paper 27, "Pet. Sur-sur-reply").

An oral hearing was held on April 19, 2023, a transcript of which appears in the record. Paper 32 ("Tr.").

B. Real Parties in Interest

Petitioner identifies Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. as the real parties-in-interest. Pet. 65. Patent Owner states that Staton Techiya, LLC is the real party in interest. Paper 31, 1.

C. Related Proceedings

Petitioner states that the '424 patent was asserted in *Staton Techiya, LLC v. Samsung Electronics Co., Ltd.*, No. 2:21-cv-00413 (E.D. Tex.), which was filed on November 5, 2021, and consolidated with *Staton Techiya, LLC v. Samsung Electronics Co., Ltd.*, Case No. 2:22-cv-00053 (E.D. Tex.), filed February 14, 2022 (the "District Court Litigation"). Pet. 65; Paper 5, 1; Paper 28, 1.

D. The '542 Patent

The '542 patent is titled "Automatic Sound Pass-Through Method and System for Earphones." Ex. 1001, code (54). The '542 patent discloses "[e]arphone systems and methods for automatically directing ambient sound to an earphone device." *Id.* at code (57). "During detected voice activity, incoming audio content is attenuated while ambient sound is increased and provided to the earphone device." *Id.* According to the '542 patent, "[u]ser voice activity is detected by analysis of at least one of an ear canal microphone signal or an ambient sound microphone signal." *Id.*

The '542 patent discloses that

> Signal processing system 206 receives an audio content (AC) signal 320 from a remote device (such as a communication device (e.g. mobile phone, earphone device 220, earphone device 222, etc.) or an audio content delivery device (e.g. music player)). Signal processing system 206 further receives ASM [Ambient Sound Microphone] signal 322 from ASM 120 (FIG. 1).
>
> A linear gain may be applied to AC signal 320 by AC gain stage 304, using gain coefficient Gain_AC, to generate a modified AC signal. In some embodiments, the gain (by gain stage 304) may be frequency dependent. A linear gain may also be applied to ASM signal 322 in gain stage 306, using gain coefficient Gain_ASM, to generate a modified ASM signal. In

> some embodiments, the gain (in gain stage 306) may be frequency dependent.

Ex. 1001, 7:13–26.

When it is determined that "the user of earphone device 100 (FIG. 1) is speaking," the gain applied to AC signal 320 is set to a low value and the gain applied to ASM signal 322 is set to a high value. *Id.* at 7:54–55, 7:58–60. When no user voice activity is detected, the gain applied to AC signal 320 is set to a high value and the gain applied to ASM signal 322 is set to a low value. *Id.* at 7:55–58. The '542 patent discloses that "[s]ignal processing system 206 may include optional VAD timer system 310" which "may provide a time period of delay (i.e., a pre-fade delay), between cessation of detected voice activity and switching of gains by gain states 304, 306 associated with the VAD off state." *Id.* at 8:1–5.

E. Illustrative Claim

Of challenged claims 1–20, claims 1, 7, and 13 are independent. For purposes of the issues raised in this proceeding, claim 1 is illustrative and is reproduced below.

> 1. [preamble] A method for passing ambient sound to an earphone device configured to be inserted in an ear canal of a user, the method comprising the steps of:
>
> [a] capturing the ambient sound from an ambient sound microphone (ASM) proximate to the earphone device to form an ASM signal;
>
> [b] receiving an audio content (AC) signal from a remote device;
>
> [c] detecting voice activity of the user of the earphone device;
>
> [d] mixing the ASM signal and the AC signal to form a mixed signal, such that, in the mixed signal, an ASM gain

> of the ASM signal is increased and an AC gain of the AC signal is decreased when the voice activity is detected;
>
> [e] detecting a cessation of the voice activity;
>
> [f] delaying modification of the ASM gain and the AC gain for a predetermined time period responsive to the detected cessation of the voice activity; and
>
> [g] directing the mixed signal to an ear canal receiver (ECR) of the earphone device.

Ex. 1001, 11:42–60.

II. DISCUSSION

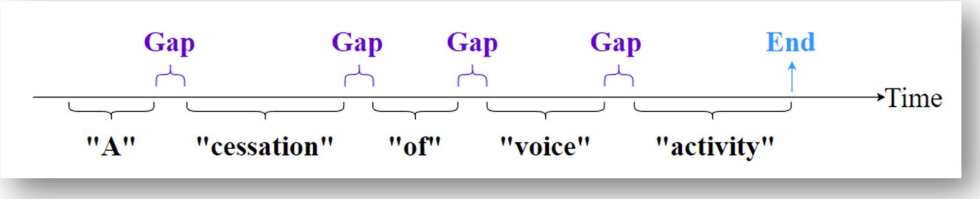
A. Claim Construction

A claim "shall be construed using the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. § 282(b)." 37 C.F.R. § 42.100(b) (2021).

1. "Detecting a Cessation of the Voice Activity"

The parties dispute the construction of the phrase "detecting a cessation of the voice activity" in claims 1 and 13. PO Resp. 18–23; Pet. Reply 2–9; PO Sur-reply 1–10. Patent Owner argues that "a cessation of the voice activity" should be construed as "an end of speech" and "does not encompass momentary pauses within speech." PO Resp. 18–19 (citing Ex. 2006 ¶ 67). Patent Owner construes "detecting a cessation of the voice activity" as "detecting an end of speech." *Id.* at 23. Petitioner responds that the plain and ordinary meaning of "detect[ing] a cessation of voice activity" is not limited to "detecting only the end of a sentence or conversation," but rather "is broad enough to also encompass detecting the end of a word within a sentence or conversation." Pet. Reply 2.

Patent Owner provides the following diagram to illustrate its argument that "momentary pauses within speech" are different from "an end of speech":



Patent Owner's figure illustrating its argument that "gaps" (momentary pauses) in speech (purple) are different than an "end of speech" (blue). PO Resp. 19.

PO Resp. 19 (citing Ex. 2006 ¶ 67).

Patent Owner argues that one of ordinary skill would have understood "a cessation of the voice activity" to mean "an end of speech" because "[a]djustment to sound signals in response to detected speech of a speaker are typically performed during the entire speech of that speaker, not during select portions of their speech." PO Resp. 19. "Otherwise," Patent Owner asserts, "the sound signal adjustments would be too frequent and too abrupt, thereby creating an unnatural speech pattern that would disturb a user's hearing experience." *Id.* (citing Ex. 2006 ¶ 68). Patent Owner also contends that, during Dr. Polish's deposition, he "distinguished 'word gaps,' i.e., the brief pauses within a question or answer[,] from a 'cessation of voice activity,' which may for example be the end of a question or answer." *Id.* at 19–20 (quoting Ex. 2007, 35:7–36:6).

Patent Owner also argues that interpreting a "cessation of voice activity" as an "[e]nd of speech" is "the only interpretation consistent with the disclosures and the goal of the '542 patent." PO Resp. 20. Specifically, Patent Owner asserts, the '542 patent seeks to reduce the isolation imposed

by sound isolating earphones by increasing the ambient sound microphone (ASM) gain and decreasing the audio content (AC) gain when user voice activity is detected so the user can hear the other person speaking. *Id.* at 21–22 (citing Ex. 1001, 1:33–36, 11:50–54, 12:24, 12:33–36; Ex. 2006 ¶ 70). When the user stops speaking, according to Patent Owner, the '542 patent restores the ASM and AC gains to their original values, but doing so too quickly would result in an increase and decrease in music volume after each spoken word. *Id.* at 22 (citing Ex. 2006 ¶ 71). By waiting until after the conversation has ended to restore the original gains using a "pre-fade delay," Patent Owner contends, this problem is overcome. *Id.* (citing Ex. 2006 ¶ 72; Ex. 1001, 8:3–4).

Petitioner responds that the plain and ordinary meaning of "detect[ing] a cessation of voice activity" is not limited to "detecting only the end of a sentence or conversation," but rather "is broad enough to also encompass detecting the end of a word within a sentence or conversation." Pet. Reply 2. Petitioner argues that "[d]etecting a user's 'voice activity' in the '542 patent refers to comparing a signal value to a threshold value." *Id.* at 3 (citing Ex. 1001, 4:1–21, Figs. 4–5, 6B). Specifically, Petitioner asserts, the '542 patent discloses that "a microphone signal level (e.g., from the ASM signal or the ECM signal) is compared to a threshold microphone value" and, if the signal level is greater than a threshold microphone value, then "voice activity is considered detected and a VAD value is set to 'on.'" *Id.* (citing Ex. 1001, 4:1–21, 7:47–50, 9:3–11, 9:61–67, 10:36–41, Figs. 4–5, 6B (step 618)). "Conversely," Petitioner contends, if the signal level "is not greater than the threshold microphone value," then "voice activity is

considered to be absent or to have ceased and a VAD value is set to 'off.'" *Id.*

Petitioner argues that the '542 patent uses the terms "cessation of the detected user voice activity" (Ex. 1001, 3:5–60, 10:46–47), "cessation of the user voice activity" (*id.* at 3:65–66, 10:54), "cessation of voice activity," (*id.* at 11:9), and "cessation of voice activity detection" (*id.* at 11:20) "interchangeably," and all these terms "refer to a state in which the signal level(s) being examined does not meet the predetermined condition and the VAD state is set to 'off.'" Pet. Reply 3–4. According to Petitioner, "[t]his comparison has nothing to do with a semantic determination of whether the user has subjectively finished speaking," and "there is no difference between a microphone signal level that is below the microphone threshold value because of the silence between words" and one that is below the threshold value because of the silence "after the user has finished speaking." *Id.* at 4–5 (citing Ex. 1025, 29:11–30:5).

Petitioner argues that its understanding of the relevant claim language "is further supported by the fact that the claims refer to 'detect[ing] a cessation of ***the*** voice activity,' the antecedent for which is 'detect[ing] voice activity of the user of the earphone device' earlier in the claim." Pet. Reply 5. Petitioner asserts that Patent Owner's construction "would have 'voice activity' mean one thing for purposes of its detection (i.e., whether a microphone signal level is greater than a threshold value), but mean a different thing for purposes of detecting whether it has ceased (i.e., a semantic determination of whether the user has finished speaking)." *Id.* Petitioner further argues that "cessation" is not a term of art, and cites dictionaries and the deposition testimony of Mr. Kleinschmidt to support its

argument that the ordinary meaning of this term "encompasses a temporary ceasing and is not limited to a permanent or final end." *Id.* (citing Ex. 1025, 22:6–13; Exs. 1026, 1027).

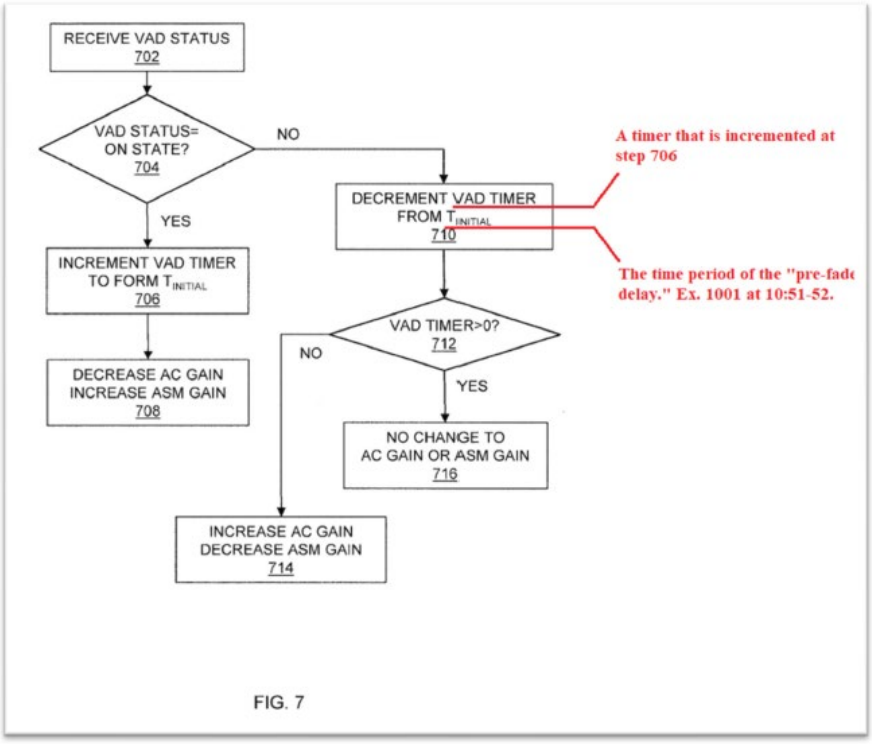
Additionally, Petitioner argues that the concepts of "detecting cessation of voice activity" and "trying to not to revert their gains to their original values before the user has stopped speaking" are "distinct." Pet. Reply 7. According to Petitioner, the first concept "is a binary determination—voice activity is either detected or it is not, and it does not matter why it is not detected." *Id.* at 7–8. The second concept, Petitioner asserts, "is addressed by providing a time period of delay between cessation of detected voice activity and reverting the gains back to their original values so as to account for natural pauses between words when speaking." *Id.* at 8. Petitioner contends that this time period of delay "is keyed off of 'cessation of detected ***voice activity***,' not the user finishing speaking" because "the system does not know whether the user intends to speak a word, a sentence, or a paragraph," and can only detect "whether the signal level of interest meets the predetermined threshold value." *Id.* at 7.

Finally, Petitioner argues that the testimony of Mr. Kleinschmidt upon which Patent Owner relies does not address why one of ordinary skill would have understood "a cessation of voice activity" to mean "an end of speech" in view of the intrinsic record, but "[at] best" only "addresses why it would be desirable not to revert the gains to their original values before the user has finished speaking." Pet. Reply 8 (citing Ex. 2006 ¶¶ 67–74). According to Petitioner, Dr. Kleinschmidt's opinion is also "at odds with his deposition testimony that the microphone signal-level value will be below the microphone signal threshold value after a pause between words just as much

as after the end of a sentence." *Id.* (citing Ex. 1025, 29:11–30:5). Finally, Petitioner characterizes Dr. Polish's testimony distinguishing word gaps from a cessation of voice activity as an "off-the-cuff response to a question directed to a legal claim construction issue that he had not had the opportunity to consider before." *Id.*

Patent Owner responds that Petitioner's "plain and ordinary meaning" arguments and Mr. Kleinschmidt's testimony that the dictionary definition of "cessation" was "fair" are not "in the context of the claims of the '542 patent." PO Sur-reply 5–6 (citing Pet. Reply 2; Ex. 1025, 21:12–22:13). Patent Owner also argues that the "pre-fade delay" discussed in the '542 patent specification "is proportional to 'continuous user voice activity (before the voice activity is ceased)'" and therefore "cannot independently determine that a user has ended their speech." *Id.* at 6. "Under Petitioner['s] interpretation of 'cessation of the voice activity,'" Patent Owner asserts, "the conclusion of every word of a user's spoken sentence would trigger the pre-fade delay" leading to "an adjustment of ASM and AC gains," which "would be too frequent and too abrupt, thereby creating an unnatural speech pattern that would disturb a user's hearing experience." *Id.* (citing PO Resp. 19).

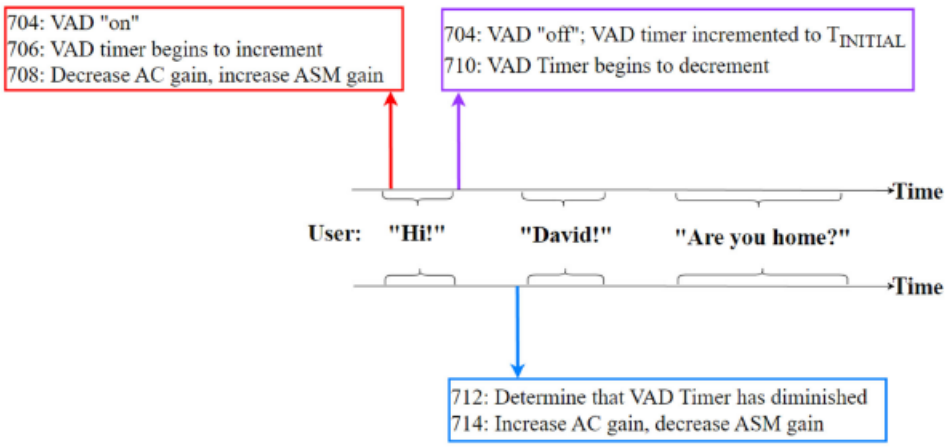
Patent Owner further argues that Figure 7 of the '542 patent supports its argument by "expressly includ[ing] the notion of detecting a cessation of voice activity and then delaying signal modifications for a period of time after such detection." PO Sur-reply 7. Patent Owner presents an annotated version of Figure 7, reproduced below.



Patent Owner's annotated version of Figure 7 of the '542 patent indicating that $T_{INITIAL}$ is "[a] timer that is incremented at step 706" and that the VAD timer in step 710 indicates "[t]he time period of the 'pre-fade delay." PO Sur-reply 8 (citing Ex. 1007, Fig. 7).

Patent Owner argues that "at step 706, following detection of user voice activity, the VAD time increments to $T_{INITIAL}$," which is the time period of the "pre-fade delay." *Id.* at 8. According to Patent Owner, $T_{INITIAL}$ has "an upper limit value, but no lower limit value," which "is consistent with the 'pre-fade delay' being proportional to 'continuous voice activity (before the voice activity is ceased).'" *Id.* at 8–9 (citing Ex. 1001, 3:59–67, 10:51–54).

Patent Owner argues that Figure 7 demonstrates that "[i]f 'cessation of the voice activity' were to include 'the end of a word within a sentence or conversation,'" then "any brief gap between words would result in an AC/ASM gain adjustment." PO Sur-reply 9. Patent Owner creates a diagram to illustrate this operation, which is reproduced below.



Patent Owner's diagram illustrating its understanding of how, in Figure 7 of the '542 patent, a brief gap between words would result in an AC/ASM gain adjustment. PO Sur-reply 9.

As shown in the Figure above, Patent Owner argues that, "upon detecting the user's voice activity at 'Hi,' the VAD timer would begin to increment, AC gain (such as the music) would decrease, and ASM gain (such as the environmental sound) would increase (red box)." *Id.* According to Patent Owner, if "cessation of the voice activity" includes "the gap between 'Hi' and 'David' (purple box), then the VAD timer would have incremented to an extremely low $T_{INITIAL}$ value" at the end of the word "Hi," when the VAD timer begins to decrement. *Id.* at 9–10. "Any slight pause following 'Hi,'" Patent Owner contends, "would therefore dim[in]ish this low $T_{INITIAL}$ value and trigger step 714 (blue box)," in which "AC gain would increase (the music becomes louder) and ASM gain would decrease (the environmental sound becomes quieter," resulting in the "unnatural speech pattern" that Patent Owner identifies. *Id.* at 10 (citing Ex. 1001, Fig. 7; PO Resp. 19).

Based on the full trial record, we agree with Petitioner that "detect[ing] a cessation of voice activity" is not limited to "detecting only the end of a sentence or conversation," but rather "is broad enough to also

encompass detecting the end of a word within a sentence or conversation." *See* Pet. Reply 2.

First, we consider the meaning of the term "voice activity" in the context of claim 1. Claim 1 recites both "detecting ***voice activity*** of the user of the earphone device" (claim limitation 1[c]) and "detecting a cessation of ***the voice*** activity" (claim limitation 1[e]). Ex. 1001 11:50, 11:55 (emphasis added). We agree with Petitioner that "detecting voice activity of the user" in limitation 1[c] refers to comparing a microphone signal value to a threshold value. *Id.* at 3. For example, the '542 patent explains that a microphone signal level from the ASM or ECM signal is compared to a threshold value, and if it is greater, then voice activity is considered detected and a VAD value is set to "on." Ex. 1001, 4:1–21, 7:47–55, 10:36–41, Figs. 4–5, 6B (step 618). Correspondingly, we agree that in the phrase "detecting a cessation of ***the voice activity***" in limitation 1[e], the word "the" before "voice activity" refers back to the voice activity detected in limitation 1[c], which provides antecedent basis. *See* Pet. Reply 4. Thus, "detecting a cessation of the voice activity" refers to detecting that the ASM or ECM microphone signal level has fallen below the threshold value, and setting the VAD value to "off." As Petitioner persuasively explains, it would not make sense for "voice activity" to have one meaning for purposes of detecting when voice activity is present, and another meaning for purposes of detecting when voice activity has ceased. *Id.* at 5; *see* Ex. 1025, 29:11–30:5 (Mr. Kleinschmidt acknowledging that a microphone signal can be below the threshold value regardless of whether there is only a gap between words as opposed to a gap after the user has finished speaking).

We next turn to the meaning of the term "cessation." We agree with Petitioner that "cessation" is not a term of art and should be given its ordinary meaning. Pet. Reply 5. Petitioner introduces the definition of "cessation" from the Merriam Webster Collegiate Dictionary as "a ***temporary*** or final ceasing (as of action)," and the definition from the New World Dictionary as "a ceasing, or stopping, either forever ***or for some time***." Exs. 1026 (emphasis added), 1027 (emphasis added). Mr. Kleinschmidt acknowledged during cross-examination that he was "not aware of a different definition of the word [cessation] in a technical sense" than the one used "in the Merriam-Webster dictionary," and that the definition of "a temporary or final ceasing (as of action)" is "a fair definition for the word 'cessation.'" Ex. 1025, 22:6–13, 23:21–24:4. Based on the intrinsic evidence, we find that "a cessation of the voice activity" covers a situation where the microphone-signal level falls below the threshold value, indicating that the voice activity has stopped, either temporarily or permanently. And, "detecting a cessation of the voice activity" involves detecting that the microphone signal level has fallen below the threshold value, either temporarily or permanently.

We do not agree with Patent Owner's argument that we should construe "detecting a cessation of the voice activity" as "detecting an end of speech." *See* PO Resp. 18–19, 23. Patent Owner appears to interpret "an end of speech" to mean the end of "the entire speech of th[e] speaker," rather than "select portions of their speech." *See* PO Resp. 19. For example, Patent Owner argues that an "end of speech" would occur at the end of a question or answer rather than during pauses within a question or answer. *Id.* at 19–20. However, we agree with Petitioner that, under this

interpretation, it is unclear how the system would know when a user's "entire speech" has ended as opposed to a pause during speech. *See* Pet. Reply 4–5. For example, a speaker could pause to think between words or in the middle of a sentence in a question or answer, even though the user has not finished their entire speech, but it is unclear how the system would know the difference. Similarly, a speaker could be talking to another person but stop to refer to something or look something up, even though the user is not finished speaking.

We also agree with Petitioner that "detecting cessation of voice activity" and determining when to revert the changes in AC and ASM gain to their original values are distinct concepts that are addressed by different limitations in claim 1. *See* Pet. Reply 7. As discussed above, "detecting a cessation of the voice activity" in claim limitation 1[e] is a binary decision based on whether the detected microphone signal(s) exceed a certain threshold. For purposes of limitation 1[e], voice activity is detected or not detected based on that determination, and it does not matter why it is not detected, or whether the user subjectively intends to continue speaking or has finished their speech. The concept of determining whether gains should be reverted to their original values is addressed not by step 1[e], but rather by step 1[f], which recites "delaying modification of the ASM gain and the AC gain for a predetermined period responsive to the detected cessation of the voice activity."

For this reason, we disagree with Patent Owner's argument that its construction of "detecting a cessation of the voice activity" is necessary because otherwise the AC and ASM gain adjustments "would be too frequent and too abrupt, thereby creating an unnatural speech pattern that

would disturb a user's hearing experience." *See* PO Resp. 19. The problem of reverting changing the AC and ASM gains too frequently is addressed by limitation 1[f], which requires delaying the modification of the ASM and AC gains "for a predetermined time period responsive to the detected cessation of the voice activity." By using a timer to delay modification of the AC and ASM gains, the system can prevent the gains from reverting to their original values too quickly, because the gains will not revert until the timer expires. By using a longer time delay, one could increase the likelihood that the gains will not be reverted to their original values prematurely, such as between brief pauses or word gaps, because the length of the timer would exceed such pauses. Thus, the problem of too quickly reverting the AC and ASM gains to their original values and creating an unnatural speech pattern need not be addressed by step 1[e].

Our conclusion in this regard is supported by Figure 7 of the '542 patent specification, which is consistent with Petitioner's interpretation of the disputed claim language. Figure 7 is reproduced below.

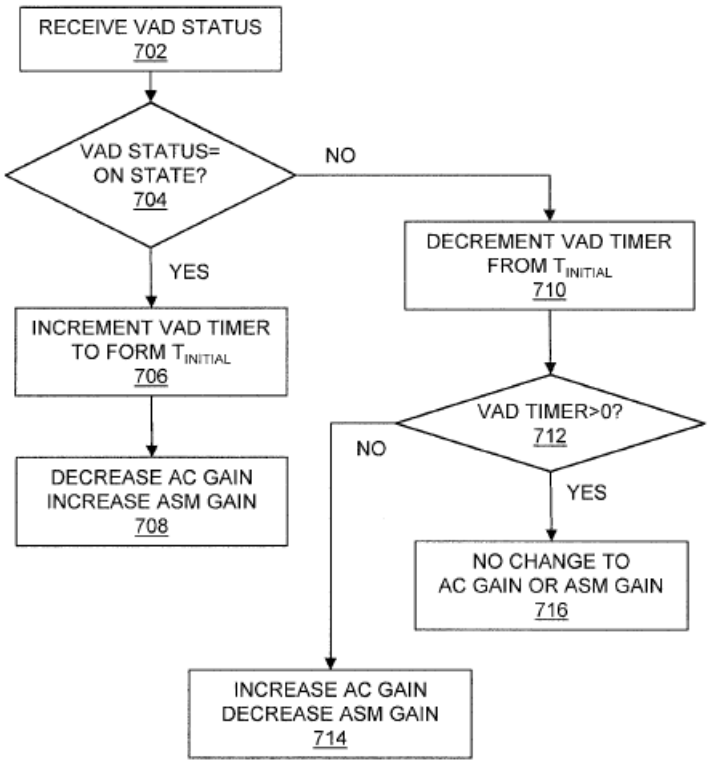


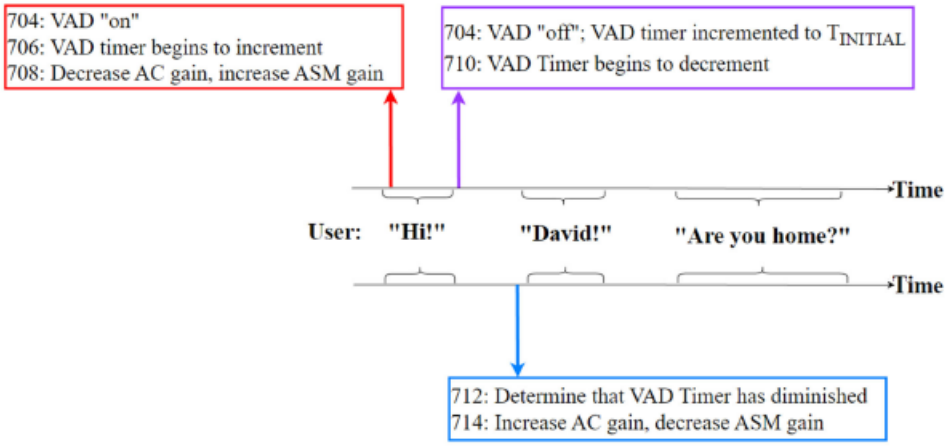
FIG. 7

Figure 7 of the '542 Patent illustrates a method for controlling AC gain and ASM gain according to an embodiment of the invention. Ex. 1001, Fig. 7, 2:38–41.

As shown in Figure 7, when the microphone detects speech and the VAD status indicates an "on" state (steps 702, 704), the system decreases AC gain and increases ASM gain (step 708). *Id.*, Fig. 7. 10:42–50, 10:58–11:2. When the speech stops (for whatever reason, whether due to a word gap or the user finishing speaking), the VAD status indicates an "off" state (steps 702, 704), and the system then uses a VAD timer to delay the reversion of the AC and ASM gains to their original values (steps 710–716). The detecting that speech has stopped and the setting of the VAD status to "off" corresponds to the step of "detecting a cessation of the voice activity"

(limitation 1[e]), and the steps of delaying reversion of the AC and ASM gains to their original values corresponds to the step of "delaying modification of the ASM gain and the AC gain for a predetermined time period responsive to the detected cessation of the voice activity" (limitation 1[f]. Indeed, Patent Owner's counsel acknowledged at the oral hearing that the "cessation of voice activity" occurs at "Decision Block 704," and that "in one embodiment," the "VAD status" is "in an off state whenever the user stops speaking." Tr. 38:16–23, 39:19–23, 41:6–11.

We do not agree with Patent Owner's argument that Figure 7 supports its construction. *See* PO Sur-reply 9–10. As noted above, Patent Owner's argument is based on the diagram below that it uses to illustrate its position.



Patent Owner's diagram illustrating its understanding of how, in Figure 7 of the '542 patent, a brief gap between words would result in an AC/ASM gain adjustment. PO Sur-reply 9.

Patent Owner's argument relies on the use of the VAD timer in Figure 7, which is incremented to form a value of $T_{INITIAL}$ when the VAD status is in an "on" state, and is decremented from the $T_{INITIAL}$ value after the VAD status switches to an "off" state. *See id.* at 9–10. Patent Owner argues that if the gap between the words "Hi" and "David" (illustrated by the red and purple arrows) is a "cessation of the voice activity," $T_{INITIAL}$ would have

only been incremented to a low value by the end of the word "Hi," resulting in a short period for the VAD timer to decrease to zero, and triggering the reversion of the AC and ASM gains before the word "David" is spoken, creating an "unnatural speech pattern." *See id.* Thus, according to Patent Owner, "cessation of the voice activity" cannot occur during such a pause.

Patent Owner's argument suffers from several infirmities. First, as explained above, Figure 7 matches up with Petitioner's understanding of the claim language rather than Patent Owner's. Specifically, the switching of the VAD status to an "off" state when the user stops speaking corresponds to the claimed "detecting a cessation of the voice activity" (limitation 1[e]) and the decrementing of the VAD timer before reverting the AC and ASM gains to their original values corresponds to the claimed "delaying modification of the ASM gain and the AC gain for a predetermined time period responsive to the detected cessation of the voice activity" (limitation 1[f]). Patent Owner appears to match the decrementing of the VAD timer to the "delaying modification" in limitation 1[f]. *See* PO Sur-reply 9–10. However, Patent Owner does not explain how Figure 7 would determine that there has been an "end of speech" ***before*** beginning to decrement the VAD timer in step 710. *See id.* As discussed above, the setting of the VAD status to "on" and "off" states is based on whether the microphone signal is above or below a threshold, not whether a user has completely finished speaking.

Patent Owner also relies on unsupported assumptions in its argument based on Figure 7. Specifically, Patent Owner appears to assume that the VAD timer would elapse and revert the gains back to their original values before the end of the gap between the words "Hi" and "David," because $T_{INITIAL}$ would start at zero and/or the VAD timer would increment too

slowly to achieve a pre-fade delay time longer than the gap between these words. *See id.* However, we see no basis for these assumptions in the '542 patent. The '542 patent states that "the time period of the 'pre-fade delay' (referred to herein as '$T_{INITIAL}$') may be ***proportional*** to a time period of continuous user voice activity" (Ex. 1001, 10:51–54 (emphasis added)), but that does not mean that the time period of the pre-fade delay is ***identical*** to the period of continuous user voice activity. For example, the pre-fade delay could be a multiple of the time period of continuous user voice activity, such as twice its length. Additionally, although the '542 patent does not expressly specify a lower bound for the pre-fade delay, that does not mean that one cannot exist. Moreover, even if Patent Owner were correct that our understanding of Figure 7 could in certain circumstances result in restoring the original AC and ASM gains between word gaps, that would only mean that the embodiment of Figure 7 would not always work in an optimal manner, and would not necessarily mean that Figure 7 cannot match up to steps 1[e] and 1[f] under Petitioner's interpretation of the claim language.

Moreover, during the oral hearing, Patent Owner's counsel was unable to identify an embodiment in the '542 patent specification that corresponds to its interpretation of the relevant claim language. Tr. 41:16–42:2. Patent Owner's counsel explained that "[t]he closest we would have . . . would be Figure 7 and our supporting explanation on Pages 8 through 10 of our sur-reply." *Id.* Patent Owner's counsel continued that its proposed interpretation of the claim language could be implemented by having "a first timer that would determine a certain amount of time has passed before there's a cessation of voice activity, and then you would have a second timer that would impose a delay after that first timer has run

down," but acknowledged that "there isn't any disclosure of these first and second timers in the specification." *Id.* at 42:6–43:6. We find that the presence of an embodiment in the specification that matches Petitioner's construction of the claim language (Figure 7) and the absence of an embodiment that matches Patent Owner's construction further supports our conclusion that Petitioner's construction is the correct one.

Finally, we turn to Patent Owner's reliance on Dr. Polish's deposition testimony, where he stated that "brief pauses in [a] sentence" are "word gaps," not "brief cessations of . . . voice activity." Ex. 2007, 35:7–36:6; *see also* PO Resp. 20. Although we agree with Patent Owner that this extrinsic evidence provides some support for its position, we have weighed it with the other evidence of record, and find that, on balance, it does not outweigh the intrinsic and extrinsic evidence discussed above that supports Petitioner's interpretation of the claim language, including the claim language itself, dictionary definitions, and the specification.

Consequently, we find that "detecting a cessation of voice activity" is broad enough to encompass detecting the end of a word within a sentence or the end of a sentence within a conversation, and is not limited to a complete ending of a person's speech.

2. The District Court Litigation

Patent Owner argues that Petitioner has taken different claim construction positions in this proceeding and in the District Court Litigation. *See* Sur-reply 2–5; Sur-sur-reply *generally*. Specifically, in this proceeding Petitioner has opposed Patent Owner's proffered construction that "detecting a cessation of the voice activity" means "detecting an end of speech." *See* Section III.A.1, *supra*. By contrast, in the District Court Litigation,

Petitioner joined Patent Owner in proposing that the same claim term means "an end of speech," and also added that this was as "opposed to [detecting] a short silence (e.g., between words)." Ex. 2009, 24, 26.

The district court adopted the proposed claim construction of "detecting a cessation of the voice activity" as "detecting an end of speech, as opposed to a short silence (e.g., between words)." *See* District Court Litigation, Docket Nos. 285 (Feb. 8, 2023, Claim Construction Order), 376 (Apr. 10, 2023, Joint Motion To Amend Claim Construction Order), and 379 (Apr. 13, 2023, Order Granting Motion No. 376). The basis for this adoption appears to be the parties' agreement, as well as Petitioner's assertion to the district court that Patent Owner's arguments in this IPR proceeding qualify as a disclaimer of claim scope. *See id.*

Patent Owner argues that we should apply the agreed constructions in the district court in this IPR proceeding "to maintain consistency of claim scope and thereby protect the integrity of the patent system." PO Sur-reply 4. In support, Patent Owner cites the Board's Consolidated Trial Practice Guide (Nov. 2019) ("CTPG"),[1] which provides that "[f]or IPR[s] . . . the AIA requires that the Office consider the effect of the regulations on the economy [and] the integrity of the patent system." CTPG 5. Patent Owner further argues that "Petitioners should not be permitted to successfully argue that disclaimer occurred in this proceeding and at the same time refuse to accept the alleged disclaimer." PO Sur-reply 5.

Petitioner argues that its agreement in the district court "was expressly and solely based on [Patent Owner's] disclaimer in this IPR," and that the District Court's decision adopting the agreed construction was also based on

[1] Available at https://www.uspto.gov/TrialPracticeGuideConsolidated.

Patent Owner's alleged disclaimer. Pet. Sur-sur-reply 1 (citing Ex. 1031, 6; Ex. 2008, 36–37). Petitioner cites *Aylus Networks, Inc. v. Apple Inc.*, 856 F.3d 1353 (Fed. Cir. 2017), as holding that a patent owner's "disavowal of claim scope in an IPR proceeding can be used against [the patent owner] in subsequent proceedings outside of the IPR." *Id.* at 2. Petitioner also cites *CUPP Computing AS v. Trend Micro Inc.*, 53 F.4th 1376 (Fed. Cir. 2022), as holding "the Board need not, and should not, apply [a patent owner's argument as a disavowal] within the IPR proceeding where the disavowal occurred." *Id.*

Upon considering the record presented to us, we disagree with Patent Owner's position. As discussed above, the parties have fully briefed the merits of the relevant claim construction dispute in this proceeding, and we have found Petitioner's proposed construction to be persuasive. *See* Section II.A.1, *supra*. We recognize that our construction is not the construction currently adopted by the district court. However, the district court's adoption of Patent Owner's construction appears to have been based on the parties' joint proposal in the District Court Litigation and Petitioner's assertion that there was a disclaimer of claim scope based on Patent Owner's argument in this proceeding. Ex. 2008, 36–37. Other than disclaimer, the parties' proposal to the district court does not offer any reasoning in support of the construction that we might impute to the district court. Ex. 1031, A-2, B-12. Even though the parties presented an agreed construction to the district court, here the parties present competing constructions and substantively address the merits of those constructions.

Patent Owner's argument that we should simply adopt the district court's claim construction to avoid inconsistent rulings is not persuasive. As

noted above, the district court relied on disclaimer based on Patent Owner's arguments in this IPR proceeding (Ex. 2008, 36–37), but the Federal Circuit has made clear that such a disclaimer does not apply to us in this proceeding. As the *CUPP* decision explains: "We now make precedential the straightforward conclusion we drew in an earlier nonprecedential opinion: '[T]he Board is not required to accept a patent owner's arguments as disclaimer when deciding the merits of those arguments.'" *CUPP*, 53 F.4th at 1383 (quoting *VirnetX v. Mangrove Partners Master Fund, Ltd.*, 778 F. App'x 897, 910 (Fed. Cir. 2019)). Therefore, we decline to apply disclaimer based on Patent Owner's arguments in this proceeding.

We also do not agree with Petitioner's argument that we should adopt the construction from the District Court Litigation "to maintain consistency of claim scope and thereby protect the integrity of the patent system." *See* PO Sur-reply 4. Patent Owner does not identify any legal principle that would require us to adopt the district court's construction, and we find that the generalized concerns of "maintain[ing] consistency of claim scope" and "protect[ing] the integrity of the patent system" are not sufficient reasons for us to adopt a construction that we find is not supported by the evidence before us. *See id.*

For the foregoing reasons, we conclude the District Court Litigation does not provide a persuasive basis to adopt Patent Owner's claim construction of "detecting a cessation of the voice activity" as "detecting an end of speech" in this proceeding.

B. Principles of Law

A claim is unpatentable under 35 U.S.C. § 103 if "the differences between the subject matter sought to be patented and the prior art are such

that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains." *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) where in evidence, objective evidence of non-obviousness.[2] *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966). When evaluating a combination of teachings, we must also "determine whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue." *KSR*, 550 U.S. at 418 (citing *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006)). Whether a combination of prior art elements would have produced a predictable result weighs in the ultimate determination of obviousness. *Id.* at 416–417.

In an *inter partes* review, the petitioner must show with particularity why each challenged claim is unpatentable. *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1363 (Fed. Cir. 2016); 37 C.F.R. § 42.104(b) (2020). The burden of persuasion never shifts to Patent Owner. *Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*, 800 F.3d 1375, 1378 (Fed. Cir. 2015).

We analyze the challenges presented in the Petition in accordance with the above-stated principles.

C. Level of Ordinary Skill in the Art

Petitioner contends that a person of ordinary skill in the art at the time of the alleged invention would have had "a bachelor's degree in electrical

[2] Patent Owner has not presented objective evidence of non-obviousness.

engineering, computer science, or a similar field and two years of experience in the design of digital audio systems and associated signal processing." Pet. 7–8 (citing Ex. 1002 ¶ 22). Petitioner further states that the person of ordinary skill "could have also obtained similar knowledge and experience through other means." *Id.* Patent Owner "applies the level of ordinary skill in the art that the Petition proposed." PO Resp. 18.

We adopt Petitioner's assessment of the level of ordinary skill in the art, which is supported by the '542 patent and the asserted prior art.

D. Ground 1B[3]: Asserted Obviousness of Claims 1–6 Based on Rosenberg in Combination With Ichimura[4]

Petitioner contends that claims 1–6 would have been obvious over Rosenberg in view of Ichimura. Pet. 10–26. Patent Owner disagrees, arguing that Petitioner has failed to prove that claims 1–6 would have been obvious over Rosenberg and Ichimura. PO Resp. 25–38.

1. Overview of Rosenberg (Ex. 1005)

Rosenberg is entitled "Ambient Sound Responsive Media Player." Ex. 1005, code (54). Rosenberg discloses "adjusting an output of a media

[3] Here, and elsewhere in the Decision, the identification of the grounds using designations such as "Ground 1B" refers to the designation of the grounds as presented in the Petition.

[4] Because, as discussed further below, we find that Petitioner has proven by a preponderance of the evidence that claims 1–6 are unpatentable over the combination of Rosenberg and Ichimura in Ground 1B, we need not decide whether Petitioner has also proven that claims 1–6 are unpatentable over Rosenberg alone in Ground 1A. *See Boston Scientific Scimed, Inc. v. Cook Gp. Inc.*, 809 F. App'x 984, 990 (Fed. Cir. 2020) ("We agree that the Board need not address issues that are not necessary to the resolution of the proceeding.").

player" by "processing the ambient audio signal to determine whether one or more characteristic forms are present within the ambient audio signal" and "reducing an output of a media player from a first volume to a second volume if the one or more characteristic forms are present within the ambient audio signal." *Id.* at code (57). According to Rosenberg, "[t]he characteristic forms may be, for example, a name or personal identifier of a user of the media player, the voice of a user of the media player, or an alarm or siren." *Id.*

More particularly, Rosenberg discloses that "the relative volume of the microphone content is automatically increased with respect to the musical media content within the mixed audio signal in response to detected ambient audio events," that is, the detected characteristic forms. *Id.* ¶ 25. According to Rosenberg, "if a first user calls the name of the media player user and then continues to speak, the routines of the present invention may be configured to perform an automatic volume reduction upon the detection of the name call as uttered by the first user and will maintain the volume reduction for at least as long as the first user's voice continues to be identified without a time gap of more than some threshold amount of time." *Id.* ¶ 51. That is, the time delay specifying how long the automatic volume reduction is maintained "is set to last for as long as the user who called the media player user's name continues to speak" and "is performed based upon the detected vocal identity of this other user." *Id.* Rosenberg further discloses that "the threshold is generally set such that if the first user speaks at a typical speaking pace, the volume reduction will be maintained until the first user finishes talking." *Id.*

Rosenberg teaches that the relative volumes are controlled "such that the musical media file is substantially louder than the ambient microphone signal as presented within the mixed audio content." *Id.* ¶ 56. In Rosenberg, "if a third party calls the name of the user of the media player, upon detection of that name being uttered, the user is presented with an audio mix of musical media and microphone data such that the user can easily hear the ambient environment as mixed with the musical media." *Id.*

2. Overview of Ichimura (Ex. 1006)

Ichimura is titled "Reproducing Device, Headphone and Reproducing Method." Ex. 1006, code (54). Ichimura discloses "headphone system 100 [that] is configured to detect that the headphone fitter has started or finished talking with the person depending on whether the headphone fitter speaks to the person, thereby performing automatic mode shifting." *Id.* ¶ 93. Ichimura's "headphone system 100 operates in the listening mode and a headphone fitter listens to the music which is output from the headphone 101." *Id.* ¶ 94. Figure 5 of Ichimura, reproduced below, depicts a timing diagram of Ichimura's listening and talking modes.

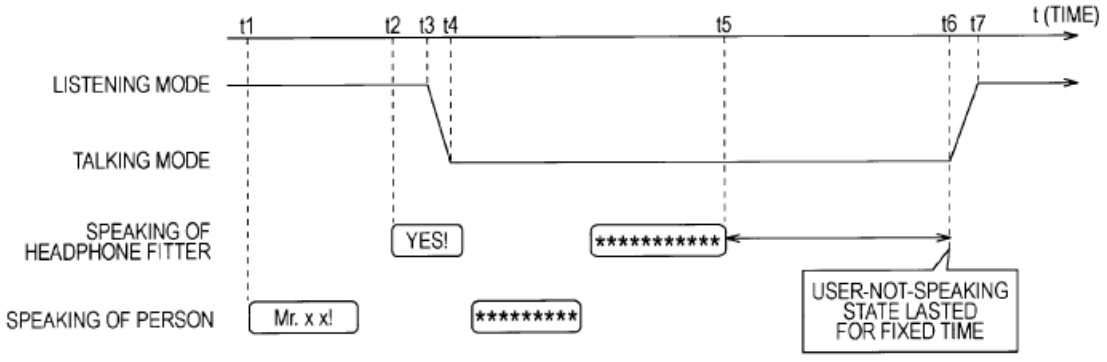


Figure 5 of Ichimura depicts "an example of a timing chart indicating timing of mode shifting from a listening mode to a talking mode and then back to the listening mode." *Id.* ¶ 24, Fig. 5.

Ichimura discloses that

> [a]t the time t2, the signal levels of the left and right mike [sic] signals become almost equal to each other and higher than the predetermined threshold value. On the basis of the above mentioned [sic] situation, the mike signal analyzing process unit 119 judges that the headphone fitter has started speaking to the person and notifies the volume and image normal position control unit 115 of the judgment.

Id. ¶ 97. According to Ichimura, "the volume and image normal position control unit 115 switches the mode from the listening mode to the talking mode at a time t3" and "performs the image normal position processing such that the normal position of the image of the reproduced sound (the music) is continuously moved from the front to the rear of the headphone fitter in a fixed time (for example, several seconds) counted from the time t3 to a time t4." *Id.* ¶ 98.

Ichimura further discloses that

> it is assumed that the headphone fitter kept talking with the person. Then, at a time t5, it is assumed that a state in which any conversation is not exchanged lasted for a fixed time after the headphone fitter has finished speaking to the person.
>
> The above mentioned [sic] situation means that a state in which the signal levels of the mike [sic] signals are lower than the threshold value lasted for a fixed time. Thus, at a time t6, the mike signal analyzing process unit 119 judges that the head phone fitter has finished talking with the person and notifies the volume and image normal position control unit 115 of this judgment.
>
> Then, at the time t6, the volume and image normal position control unit 115 switches the mode from the talking mode to the listening mode. In the above mentioned [sic] case, the volume and image normal position control unit 115 performs the image normal position processing so as to continuously move the normal position of the image of the reproduced sound from the rear to the front of the headphone

> fitter in a fixed time (for example, several seconds) counted from the time t6 to a time t7.

Id. ¶¶ 100–102.

3. Analysis of Independent Claim 1

a) 1[preamble]: "[a] method for passing ambient sound to an earphone device configured to be inserted into an ear canal of a user, the method comprising the steps of:"

Petitioner argues that, to the extent the preamble is limiting, Rosenberg discloses or suggests it. Pet. 9 (citing Ex. 1002 ¶¶ 75–86). Petitioner asserts that Rosenberg describes "a microphone 95A which is used for the detection of sound signals from the user's local ambient environment," and a media player 100 selectively mixing musical audio content with the ambient sounds so that "the user can listen to musical media content *in audio combination with ambient audio signals from the local environment.*" *Id.* (citing Ex. 1005 ¶¶ 25–26, 39; Ex. 1002 ¶ 75). According to Petitioner, Rosenberg also describes passing the audio combination to "the ears of a user" via "headphones (or other similar personalized audio presentation units that display audio content to the ears of a user) 95B," including "headsets and ear pieces." *Id.* (citing Ex. 1005 ¶¶ 20, 40; Ex. 1002 ¶¶ 76–77). Petitioner contends that one of ordinary skill "would have understood Rosenberg's description of ear pieces to be sufficiently broad to encompass 'earphone devices configured to be inserted in an ear canal of a user,' as this was a commonly known configuration for ear pieces." *Id.* at 9–10 (citing Ex. 1002 ¶ 78).

Petitioner argues that, to the extent Rosenberg's description of "ear pieces" does not describe or suggest an earphone device configured to be inserted in an ear canal, one of ordinary skill "would have been motivated to

design *Rosenberg*'s 'ear pieces' to be inserted into a user's ear canal at least because such earphones were known to be advantageously compact, portable, and suitable for isolating the listener from ambient noise, which was known to provide improved playback quality and hearing protection." Pet. 10 (citing Ex. 1002 ¶ 84). "Such a configuration of Rosenberg's ear pieces (i.e., in the ear canal of a user)," Petitioner asserts, would have been obvious to one of ordinary skill "because it would have amounted to nothing more than use of a known technique (earphones that include a portion inserted into a user's ear canal) to improve similar devices (Rosenberg's ear pieces) in the same way (including a portion inserted into a user's ear canal)." *Id*. Such a configuration," Petitioner argues, would also have been obvious to one of ordinary skill because "it would have amounted to nothing more than applying a known technique (earphones that include a portion inserted into a user's ear canal) to a known device (Rosenberg's ear pieces) ready for improvement to yield predictable results (isolating the listener from ambient noise, which was known to provide improved playback quality and hearing protection)." *Id.* at 10–11 (citing Ex. 1002 ¶ 85).

Accordingly, Petitioner argues, "Rosenberg discloses or suggests displaying ambient sounds from a local environment ('passing ambient sound') to a user via ear pieces that are placed in a user's ear canal ('to an earphone device configured to be inserted in an ear canal of a user')." Pet. 11 (citing Ex. 1002 ¶ 86).

Patent Owner does not present arguments regarding the preamble. PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that the preamble would have been obvious over the prior art.[5]

b) 1[a]: "capturing the ambient sound from an ambient sound microphone (ASM) proximate to the earphone device to form an ASM signal;"

Petitioner argues that Rosenberg discloses or suggests this feature. Pet. 11 (citing Ex. 1002 ¶¶ 87–89). Petitioner asserts that Rosenberg discloses "a microphone 95A which is used for the detection of sound signals from the user's local ambient environment," and further provides that "[a]mbient sound signals from microphone 95A are generally captured as analog audio signals and converted to digital form by an analog to digital converter or other similar component and/or process." *Id.* (citing Ex. 1005 ¶ 39). According to Petitioner, Rosenberg explains that the resulting "digital signal representing the ambient audio content captured by microphone 95A" is then provided to the media player's processor. *Id.* (citing Ex. 1002 ¶ 87) (emphasis omitted). Additionally, Petitioner contends, Rosenberg discloses that "[i]n some embodiments, *the microphone 95A is local to the headphones* or other head-worn component of the user," which one of ordinary skill would have understood to comprise a microphone proximate the headphones or other head-worn component of the user, such as ear pieces. *Id.* at 11–12 (citing Ex. 1005 ¶ 39; Ex. 1002 ¶ 88).

"Accordingly," Petitioner contends, "Rosenberg discloses or suggests capturing ambient sound from microphone 95A ('capturing ambient sound from an ambient sound microphone (ASM)') local to the headphones or ear

[5] Because we are persuaded that Petitioner has shown that Rosenberg teaches the subject matter recited in the preamble, we need not decide whether the preamble is limiting.

pieces ('proximate to the earphone device') to form a digital signal representing the ambient audio content captured by microphone 95A ('to form an ASM signal')." Pet. 12 (citing Ex. 1002 ¶ 89).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

c) 1[b]: "receiving an audio content (AC) signal from a remote device;"

Petitioner argues that Rosenberg discloses or suggests this feature. Pet. 12 (citing Ex. 1002 ¶¶ 90–92). Petitioner asserts that Rosenberg discloses a media player 100 "configured for the playback and recording of digital media," and headphones or earpieces that are operatively connected with the media player through "wired or wireless connections," such as, for example, "a Bluetooth communication link." *Id.* (citing Ex. 1005 ¶¶ 38, 40). According to Petitioner, one of ordinary skill "would have understood Rosenberg's media player connected via a Bluetooth communication link to comprise 'a remote device.'" *Id.* (citing Ex. 1002 ¶ 91). "Indeed," Petitioner contends, "this is consistent with how the '542 patent describes this feature, whereby an audio content delivery device (e.g., a portable media player) or a voice communication device (e.g., a mobile phone) with which the earphone device communicates via a wired or a wireless connection is referred to as a 'remote device.'" *Id.* at 12–13 (citing Ex. 1002 ¶ 91).

"Accordingly," Petitioner argues, "Rosenberg discloses or suggests the headphones receiving an audio representation of media content ('an audio content (AC) signal') output by a media player via a Bluetooth connection link ('from a remote device')." Pet. 13 (citing Ex. 1002 ¶ 92).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

d) 1[c]: "detecting voice activity of the user of the earphone device;"

Petitioner argues that Rosenberg discloses or suggests this feature. Pet. 13 (citing Ex. 1002 ¶¶ 93–98). Petitioner asserts that Rosenberg discloses a microphone 95A for capturing sounds from the user's local ambient environment as an ambient audio signal, and processing that signal to detect the presence of a "characteristic form," also called a "detected ambient audio event." *Id.* (citing Ex. 1005 ¶¶ 25, 39, 47; Ex. 1002 ¶¶ 93-94) (citations omitted). According to Petitioner, "[t]hree examples of these forms (events) are: '(A) the utterance of the media player user's name . . . (B) the unique voice of the media player user . . ., and/or (C) the sound of an alarm and/or siren and/or other similar emergency related alert sound . . . .'" *Id.* (citing Ex. 1005 ¶ 46). One of ordinary skill, Petitioner contends, "would have understood that the 'media player user' is the user of the headphones/ear pieces ('earphone device')." *Id.*

"Accordingly," Petitioner argues, "Rosenberg discloses or suggests detecting the unique voice of the media player user ('detecting voice activity of the user of the earphone device')." Pet. 14 (citing Ex. 1002 ¶ 98).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

e) 1[d]: "mixing the ASM signal and the AC signal to form a mixed signal, such that, in the mixed signal, an ASM gain of the ASM signal is increased and an AC gain of the AC signal is decreased when the voice activity is detected;"

Petitioner argues that Rosenberg discloses or suggests this feature. Pet. 14 (citing Ex. 1002 ¶¶ 99–107). "For example," Petitioner asserts, "Rosenberg provides that 'the media player is operative to mix musical audio content derived from a stored media file with ambient audio content captured from a microphone local to the user,'" and "[t]he 'two separate audio signals' are mixed 'into a single audio stream' (a mixed signal)." *Id.* (citing Ex. 1005 ¶¶ 25, 54–58; Ex. 1002 ¶ 99). According to Petitioner, "Rosenberg further describes selectively adjusting 'the relative volume balance of the mixed signal (i.e., the relative volume of the musical media content and the ambient microphone content)' in response to detected ambient audio events such as 'the detection of the media player's own voice within the ambient audio signal.'" *Id.* (citing Ex. 1002 ¶ 55). Additionally, Petitioner contends, Rosenberg "discloses mixing the volume level by gradually decreasing the musical audio content to substantially zero while gradually increasing the ambient audio content to the prior music volume level." *Id.* (citing Ex. 1005 ¶ 58; Ex. 1002 ¶ 100).

Petitioner argues that one of ordinary skill "would have known that the adjustment of gain is a particularly well-known technique for implementing the volume level adjustments of Rosenberg, particularly in view of Rosenberg's cross-fading feature." Pet. 14–15 (citing Ex. 1002 ¶¶ 101–105). One of ordinary skill reading Rosenberg, according to Petitioner, "would have been motivated to use the well-known technique of applying a gain to Rosenberg's ambient and media audio signals to accomplish the mixing of the 'relative volume' of the two signals in

Rosenberg into the 'single audio stream' heard by the user." *Id.* at 15 (citing Ex. 1002 ¶ 106). "Indeed," Petitioner asserts, "applying gain to adjust the level of signals to be mixed together was a common way to mix signals at the time of the purported invention," and one of ordinary skill "would have been able to do so with a reasonable expectation of success using well known techniques (e.g., off-the-shelf amplifiers and mixers) for their intended purposes with predictable results." *Id.* (citing Ex. 1002 ¶ 106).

"Accordingly," Petitioner argues, "Rosenberg discloses or suggests adjusting a mixed signal such that a volume level of the ambient audio signal ('an ASM gain of the ASM signal') is increased, and a volume level of the media audio signal ('an AC gain of the AC signal') is decreased in response to detection of the media player user's own voice within an ambient audio signal ('when the voice activity is detected')." *Id.* (citing Ex. 1002 ¶ 107).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

f) 1[e]: "detecting a cessation of the voice activity;"
1[f]: "delaying modification of the ASM gain and the AC gain for a predetermined time period responsive to the detected cessation of the voice activity; and"

Petitioner argues that Rosenberg in view of Ichimura renders these limitations obvious. Pet. 28 (citing Ex. 1002 ¶¶ 142–153). Petitioner asserts that "Ichimura is directed to a headphone device that shifts between a talking and listening mode based on whether the user is speaking." *Id.* (citing Ex. 1006 ¶¶ 78, 84; Ex. 1002 ¶ 143). According to Petitioner, "[t]he device receives a digital audio signal from a media player and contains two

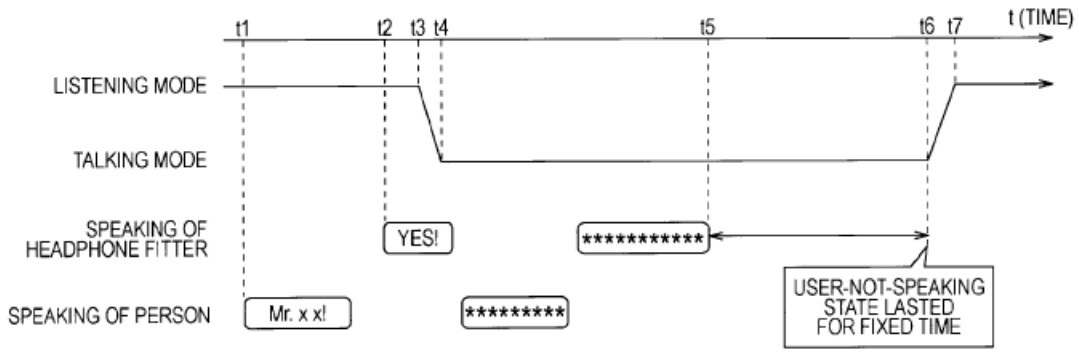
microphones for picking up external sound." *Id.* (citing Ex. 1006 ¶¶ 46–47). "In general," Petitioner contends, "the device operates in the 'listening mode,' in which the audio (e.g., music) is played through the two speakers of the headphones." *Id.* at 28–30 (citing Ex. 1006 ¶¶ 55, 59). "In listening mode," Petitioner contends, "the device generates a signal which is out of phase to the external sound picked up by the microphones," which "is mixed with the audio signal and played back to the user, resulting in a noise cancelling effect and clear tone quality of the played back audio." *Id.* at 29 (citing Ex. 1006 ¶¶ 61–62; Ex. 1002 ¶ 144).

Petitioner argues that "Ichimura further discloses performing a microphone signal analysis process on the external (or 'ambient') audio signal to determine whether the user is talking." Pet. 29 (citing Ex. 1006 ¶ 65). "When the analysis indicates that the user has started talking," Petitioner asserts, "the device switches to 'talking mode,'" which "turn[s] off the noise-cancelling function, and reproduc[es] the sound picked up by the microphones to the speakers." *Id.* (citing Ex. 1006 ¶¶ 66–69). "While Ichimura's device still plays the audio signal," according to Petitioner, "the music is output from the loudspeakers which are installed at the rear of him." *Id.* (citing Ex. 1006 ¶ 75). "Then," Petitioner contends, "when the signal analysis indicates that the user has finished talking for a certain amount of time, Ichimura describes turning the noise-cancelling back on and playing the audio signal as if the loudspeakers were 'installed in the front' of the user." *Id.* (citing Ex. 1006 ¶¶ 83, 91–92, 100–0102, 118–119, Figs. 5, 6; Ex. 1002 ¶ 145).

Thus, Petitioner argues, Ichimura "describes modifying external audio signals and digital voice signals when a state in which the headphone fitter

does not speak ('responsive to detected cessation of voice activity') has lasted for a fixed time period ('delaying modification of the ambient and audio signals for a predetermined time period')." Pet. 29–30 (citing Ex. 1006 ¶ 33). According to Petitioner, this feature is shown in Ichimura's Figure 5, reproduced below.

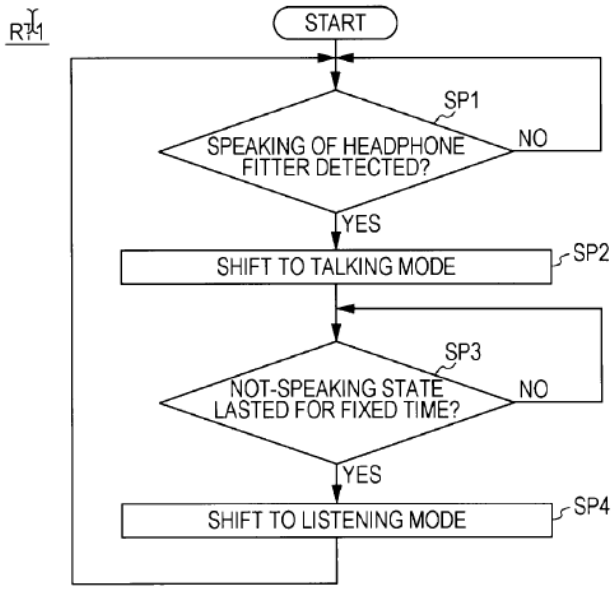
FIG. 5



Petitioner argues that Figure 5 discloses "a time delay between time t5 (when headphone fitter (or wearer) has finished talking) and time t6 (when the mode is switched from talking mode to listening mode)." *Id.* at 30.

Petitioner argues that this feature is also shown in Ichimura's Figure 6, reproduced below.

FIG. 6



As shown in Figure 6, the system first detects if the headphone fitter is speaking (step SP1). Ex. 1006, Fig. 6. If so, the system shifts to talking mode (step SP2). *Id.* The system then detects if a non-speaking state has lasted for a fixed time (step SP3), and if so, it shifts to listening mode (step SP4). *Id.*

> Petitioner argues that one of ordinary skill
>
> would have been motivated to modify Rosenberg in view of Ichimura to detect when Rosenberg's media player user has stopped talking as taught by Ichimura ("detecting a cessation of the voice activity") and to return the relative volume levels of the ambient audio signal and the music audio signal to their levels before the media player user's voice was detected as taught by Rosenberg ("modification of the ASM gain and the AC gain") only after the media player user has stopped talking for a predetermined time period as taught by Ichimura ("delaying . . . for a predetermined time period responsive to the detected cessation of the voice activity").

Pet. 31 (citing Ex. 1002 ¶ 147).

Petitioner further argues that one of ordinary skill would have been motivated to combine Rosenberg and Ichimura because both "are directed to techniques for protecting a user who is listening to music with headphones from ambient sounds except when it would be desirable to hear such sounds (e.g., when the user is engaged in a conversation with another person)." Pet. 32 (citing Ex. 1002 ¶ 148). "Whereas Rosenberg discloses using the first user's voice (i.e., the person to whom the media player user is speaking) to detect whether the media player user (i.e., the person wearing the ear pieces) is engaged in a conversation," Petitioner asserts, "Ichimura discloses using the headphone user's voice to determine whether the headphone user is engaged in conversation with another person." *Id.* (citing Ex. 1005 ¶¶ 50–51; Ex. 1006 ¶¶ 65–67) (citation omitted). "Both references," according to Petitioner, "disclose or suggest waiting a predetermined amount of time after detecting that the relevant voice has stopped talking before returning to the pre-conversation volume levels." *Id.* (citing Ex. 1005 ¶¶ 50–51; Ex. 1006 ¶¶ 91, 100; Ex. 1002 ¶ 148). One of ordinary skill, Petitioner contends, "would have recognized the benefit of using Ichimura's technique to supplement Rosenberg's system to determine that a conversation has ended by detecting that both the media player user and the first user have stopped talking, and return the mix volume to pre-event levels after both the media player user and the first user have stopped talking for a predetermined amount of time." *Id.* (citing Ex. 1002 ¶ 149).

Thus, Petitioner contends,

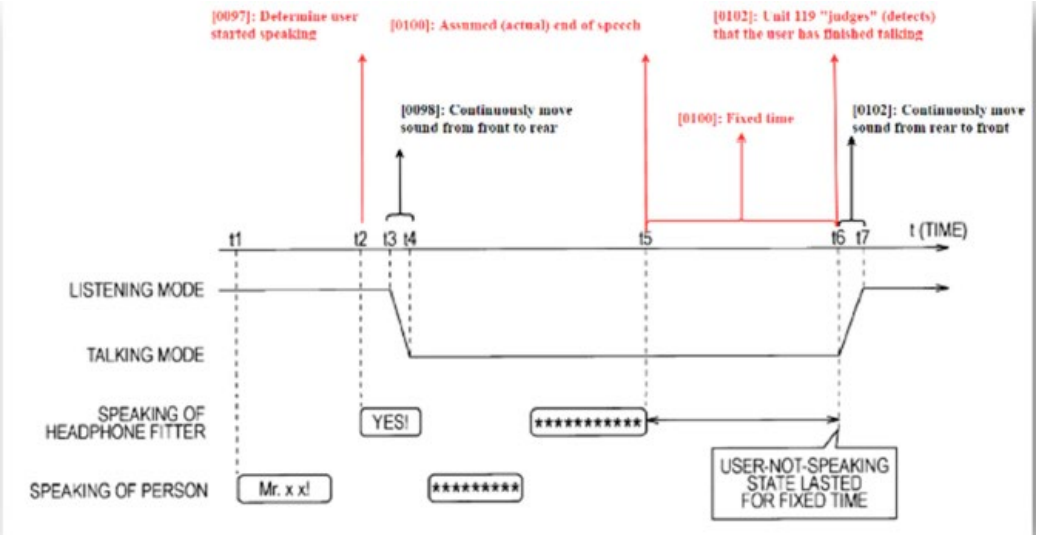
> [c]ombining the teachings of Rosenberg and Ichimura would have been obvious to a [person of ordinary skill in the art] because the combination would have amounted to nothing more than applying a known technique (Ichimura's technique of using the headphone user's voice to detect when a conversation

> has begun and ended for a fixed time period) to a known device (Rosenberg's technique of using the first user's voice to detect when a conversation has begun and ended) ready for improvement (to allow Rosenberg to use both techniques) to yield predictable results (enable Rosenberg to more accurately predict when a conversation has ended by accounting for both the media player user's voice and the first user's voice).

Pet. 32–33 (citing Ex. 1002 ¶ 150).

Alternatively, Petitioner argues, one of ordinary skill would have "recognized that Ichimura's technique could be used as a replacement in Rosenberg's system." Pet. 33 (citing Ex. 1002 ¶ 151). "In other words," Petitioner asserts, one of ordinary skill "would have recognized that Ichimura's technique could be used to modify Rosenberg's system to determine that a conversation has ended by detecting that the media player user has stopped talking, and returning volume levels to the pre-event levels after the media player user has stopped talking for a predetermined amount of time." *Id.* According to Petitioner, this alternative combination would have been obvious to one of ordinary skill "because it would have amounted to nothing more than a simple substitution of one known element (Ichimura's technique of using the headphone user's voice to detect when a conversation has begun and ended) for another (Rosenberg's technique of using the first user's voice to detect when a conversation has begun and ended) to obtain predictable results." *Id.* (citing Ex. 1002 ¶ 152).

Patent Owner responds that Ichimura does not delay sound modification "responsive to the detected end of speech," but instead "begins the process of reversion to pre-speech levels upon a detection of an end of speech without delay." PO Resp. 34. Patent Owner illustrates its argument using an annotated version of Ichimura's Figure 5, reproduced below.



Patent Owner's annotated version of Ichimura's Figure 5 illustrating Patent Owner's argument that Ichimura does not delay reversion to pre-speech levels after an end of speech. PO Resp. 35.

Referring to its annotated Figure 5, Patent Owner argues that "the actual end of speech occurs" at time t5, at which, Ichimura discloses, "it is assumed that a state in which any conversation is not exchanged lasted for a fixed time after the headphone fitter has finished speaking to the person." *Id.* at 35 (citing Ex. 1006 ¶ 100). Patent Owner asserts that in this sentence "the word 'assumed' should not be confused with the system's detection or determination of end of speech," because "Ichimura uses this word throughout its description to describe facts, not the system's determinations." *Id.* (citing Ex. 1006 ¶¶ 94–96). By contrast, according to Patent Owner, "Ichimura describes the system's detection using the word 'judges.'" *Id.* (citing Ex. 1006 ¶¶ 32, 33, 37, 39, 40). At time t6, Patent Owner argues, "the detected end of speech occurs." *Id.* at 36.

"The 'fixed time' between t5 and t6," Patent Owner argues, "does not correspond to 'delaying modification' because it does not occur 'responsive to the detected cessation of the voice activity,' as claim 1 of the '542 patent requires," but is instead "a time used to detect when speech has stopped."

PO Resp. 36. This "fixed time," Patent Owner continues, "is not a time delay added to the time when it has been determined that speech has stopped," otherwise "there would be a time delay between t6 and t7 where there would be no change in the modes after it has been determined that there has been a cessation of voice activity." *Id.* (citing Ex. 2006 ¶ 98). "And," according to Patent Owner, "at the detected end of speech, Ichimura departs from claim 1's limitation" because, at time t6, Ichimura's "unit 119 immediately 'switches the mode from the talking mode to the listening mode' by 'continuously mov[ing] the [sound position] from the rear to the front . . . in a fixed time (for example, several seconds) counted from the time t6 to a time t7.'" *Id.* at 37 (citing Ex. 1006 ¶ 101). "In other words," Petitioner contends, "upon detecting end of speech, Ichimura expressly discloses sound modification (talking mode) is no longer maintained," which "directly contradicts" claim 1's language "recit[ing] ***delaying*** sound modification 'responsive to the detected end of speech.'" *Id.*

"Additionally," Patent Owner argues, "Ichimura separates the ambient sound image from the music image and specifically states that they should not be mixed." PO Resp. 37. Patent Owner quotes the following portion of Ichimura:

> *[I]f the image of the music which is* being reproduced is positioned in the front of the headphone fitter, the image of the music which is being reproduced will be *mixed with the image of the voice* of the person and *it will become difficult for the headphone fitter to catch the voice of the person***.**
>
> Accordingly, the headphone system 100 is configured such that the normal position of the image of the music which is being reproduced is moved to the rear of the headphone fitter while he is talking with the person such that the image of the voice of the person who talks with the headphone fitter is **not mixed** with the image of the music which is being reproduced.

Id. at 37 (citing Ex. 1006 ¶¶ 72, 73). "Thus," Patent Owner contends, "Ichimura teaches away from directly mixing the ASM signal with AC signal forming a mixed signal." *Id.* at 38. To the contrary, according to Patent Owner, "Ichimura specifically separates the signals into an ASM image (image of the voice) and an AC image (image of the music), both of which can be sent to the speaker of the headphones, and specifies not mixing them." *Id.* at 38.

Petitioner responds that Patent Owner's argument that Ichimura "begins the process of reversion to pre-speech levels upon a detection of an end of speech without delay" is "based on the incorrect understanding that, at time t5 (Ichimura, Fig. 5), Ichimura does not detect that the user has stopped talking," and that the "fixed time" between t5 and t6 "is not a time delay added to the time when it has been ***determined*** that speech has stopped." Pet. Reply 14 (citing PO Resp. 34–36; Ex. 1006, Fig. 5) (citations omitted). Petitioner argues that Patent Owner's assertion is wrong because "the fact that Ichimura's system may make a determination at t6 is not inconsistent with Ichimura also ***detecting*** the cessation of voice activity at t5." *Id.* at 15. Petitioner asserts that Ichimura "clearly discloses" a "fixed time in depicting 'USER-NOT-SPEAKING STATE LASTED FOR FIXED TIME' between t5 and t6 of Figure 5.'" *Id.* According to Petitioner, "[i]t would have been impossible for Ichimura to determine at time t6 whether the 'USER-NOT-SPEAKING STATE' had lasted for a fixed time unless at t5 Ichimura had ***detected*** that the user had in fact stopped speaking." *Id.* Petitioner argues that Patent Owner "fails to engage with this aspect of Ichimura's disclosure," and that at his deposition, Mr. Kleinschmidt confirmed Petitioner's understanding of Ichimura that "[a]t t5, the system

has detected a cessation of words, of voice." Pet. Reply 15–16 (citing Ex. 1025, 80:2–6). "Thus," Petitioner asserts, "Ichimura works just as the '542 patent—it detects that the user has stopped talking at t5 ('detecting a cessation of the voice activity') and returns the sound to its pre-event state when the 'user not speaking state' has lasted for a fixed time (t6) ('delaying modification . . .')." *Id.* at 16.

Additionally, Petitioner disagrees with Patent Owner's interpretation of how Ichimura uses the terms "assumed" and "judges" because "Ichimura itself does not define these terms" and, in any event, "Ichimura would not be able to determine that a 'USER-NOT-SPEAKING STATE LASTED FOR FIXED TIME' at t6 unless it had detected that the voice activity had ceased by time t5." Pet. Reply 16.

Finally, Petitioner asserts that Patent Owner's argument that "Ichimura teaches away from mixing signals" "misses the mark" because Petitioner is "combining with Rosenberg Ichimura's teachings directed to waiting a predetermined time after detecting the user has stopped talking before returning to the original state," not "Ichimura's teachings regarding what adjustments to make when a conversation is detected." Pet. Reply 16–17 (citing Pet. 31–33; Ex. 1002 ¶¶ 147–153). Petitioner contends that "[n]othing in the record" suggests or discloses "that the aspects of Ichimura on which Petitioner[] rel[ies] are incompatible with Rosenberg." *Id.* at 17.

Patent Owner responds that "[t]he '542 patent and Ichimura treat end of speech differently," because "[i]n the '542 patent, first a cessation (i.e., and end of speech is detected), then a pre-fade delay is employed," while in Ichimura "the determination that the user finished speaking at t6 results in the immediate transition from 'Talking Mode' to 'Listening Mode.'" PO

Sur-reply 15 (citing Ex. 1001, 11:50–54; Pet. Reply 7–8). Patent Owner argues that "the parties agreed that 'a cessation of the voice activity' means 'an end of speech'" and that this construction was adopted by the district court. *Id.* "Under these constructions," Patent Owner asserts, the detecting "occurs, if at all, at time t6 in Ichimura, where "the mike signal analyzing process 119 judges that the headphone fitter ***has finished talking with the person***." *Id.* at 15–16 (citing Ex. 1006 ¶ 101). "[U]pon such detection," according to Patent Owner, "Ichimura ***immediately*** (i.e., without delay) reverts from the 'talking mode' to the 'listening mode.'" *Id.* at 16.

Based on the full trial record, we find that Petitioner has sufficiently proven that these limitations would have been obvious over the combination of Rosenberg and Ichimura. The parties' dispute centers mostly on the proper construction of the term "detecting a cessation of the voice activity," which was construed above in Section II.A. For limitations 1[d] and 1[f], Patent Owner's argument is based on its construction of "detecting a cessation of the voice activity" as meaning "detecting an end of speech," which would not cover detecting a pause between words. *See, e.g.*, PO Resp. 34 (arguing that "Ichimura's sound modification is not delayed responsive to the detected end of speech"), 36 (arguing that "[a]t t6, the detected end of speech occurs"); PO Sur-reply 15 (arguing that, under Patent Owner's construction of "a cessation of the voice activity" as "an end of speech," the detecting "occurs, if at all, at time t6"). However, for the reasons discussed in Section II.A above, we have rejected Patent Owner's proposed construction, and we have determined that "detecting a cessation of the voice activity" can include detecting pauses between words and does not require that the user has ended their speech.

Based on our construction of "detecting a cessation of the voice activity," we agree with Petitioner that limitations 1[e] and 1[f] are disclosed by the Rosenberg-Ichimura combination. Specifically, we agree that Ichimura detects a cessation of the voice activity at time t5 in Figure 5, where the system enters "a state in which the signal levels of the mike signals are lower than the threshold value," indicating that "conversation is not exchanged." Ex. 1006 ¶¶ 100–110. Ichimura also describes this operation in paragraph 91, which explains that

> [e]ven after mode shifting [to the talking mode], the mike signal analyzing unit 119 continuously performs that analyzing process. That is, the mike signal analyzing process unit 119 keeps monitoring the signal levels of the left and right mike signals. Then, when a state in which the left and right signal levels are lower than the predetermined threshold value, that is, the headphone fitter does not speak to the person lasts for a predetermined time (for example, several seconds), the mike signal analyzing process unit 119 judges that the headphone fitter has finished talking with the person.

Ex. 1006 ¶ 91. Ichimura goes on to explain that, after this judgment, "the volume and image normal position control unit 115 shifts the mode from the talking mode back to the listening mode." *Id.* ¶ 92.

Based on this discussion, we understand Ichimura to disclose that, after switching to the talking mode, the system continuously monitors the microphone signals to detect whether they drop below the predetermined threshold values, which indicates that speech has stopped. In Figure 5, the detection that the microphone signals have dropped below the predetermined threshold values occurs at time t5. At time t5, the system starts a timer to determine whether the microphone signals stay below the predetermined threshold for a fixed time, which Ichimura identifies as a "USER-NOT-SPEAKING STATE." Ex. 1006, Fig. 5, Fig. 6 (step SP3), ¶¶ 100–102. If

the microphone signals remain below the predetermined thresholds for this "fixed time," the system "judges that the headphone fitter has finished talking with the person," and shifts back to listening mode. *Id.* ¶ 91, Figs. 5, 6. We find that the step of determining at t5 that the microphone signals are below the predetermined threshold matches up to "detecting a cessation of the voice activity" in limitation 1[e], and the step of delaying switching to listening mode from t5 to t6 "delay[s] modification . . . for a predetermined period responsive to the detected cessation of the voice activity" as recited in limitation 1[f].[6]

Our determination is supported by the testimony of Dr. Polish that Ichimura's Figure 5 shows "a time delay between t5 (when headphone fitter (or wearer) has finished talking) and time t6 (when the mode is switched from talking mode to listening mode)." Ex. 1002 ¶ 146. It is further supported by the following testimony from Mr. Kleinschmidt, which we rely on and find credible:

> Q: In Ichimura Figure 5 you see in the middle of the pages between t4 and t5 two sets of asterisks; right?
>
> A: I see, yes.
>
> Q: The first set of asterisks towards the bottom of the page corresponds to the person who's speaking to the individual wearing the headset; right?
>
> A: That's correct.
>
> Q: Okay. And then there's another set of asterisks that corresponds to the speaking of the person who's wearing the headset; right?
>
> A: That's correct. Identified as "headphone fitter."

[6] Limitation 1[f] states that what is being modified is "the ASM gain and the AC gain," which is different from what is modified in Ichimura. That difference is addressed further below.

> . . . .
>
> Q: And that ends at t5; right? Per the diagram?
>
> . . . .
>
> A: Okay. That's correct. So, yes, I'd suggest those – that set of asterisks is – at t5 when the headphone fitter has stopped speaking.
>
> Q: Okay. And, also, there's no speech activity from the person who is speaking to the headphone wearer, the 'speaking of' person; right? As of t5?
>
> A: I believe that's correct.
>
> . . . .
>
> Q: T6 corresponds to the Ichimura system identifying that the individual wearing the headphone has stopped speaking because no speech is identified in the ambient signal for a fixed period of time; correct?
>
> A: Correct.
>
> A: So at t6 the Ichimura system determines that the user is done speaking because the user hasn't spoken for some fixed period of time; right?
>
> A: Correct
>
> . . . .
>
> Q: At t5, the Ichimura system detects that the voice activity has stopped; there's no more words; there's just silence; right?
>
> A: At t5 the system has detected a cessation of words, of voice.
>
> Q: And at time t6 the system has also measured the fixed period in which there has been an absence of voice in the ambient signal; right?
>
> A: T6 identifies – identifies the end of the fixed period of time when there is no voice signal.

Ex. 1025, 76:15–80:15

We also disagree with Patent Owner's argument that one of ordinary skill would not have modified Rosenberg with Ichimura because "Ichimura

separates the ambient sound image from the music image and specifically states that they should not be mixed." *See* PO Resp. 37–38 (citing Ex. 1006 ¶¶ 72, 73). As discussed in Section II.D.3(e) above, Petitioner relies on Rosenberg for limitation 1[d] requiring "mixing the ASM signal and the AC signal to form a mixed signal." Petitioner relies on Ichimura to teach "waiting a predetermined amount of time after detecting the user has stopped talking before returning to the original state," not "Ichimura's teachings regarding what adjustments to make when a conversation is detected." *See* Pet. Reply 16–17. Therefore, the fact that Ichimura does not disclose mixing the AC and ASM signals does not defeat Petitioner's argument based on the combination of Rosenberg and Ishimura.

Additionally, we disagree with Patent Owner that Ichimura "teaches away" from mixing the ASM and AC signals to form a mixed signal. *See* PO Resp. 37–38. As the Federal Circuit has explained, "[a] reference may be said to teach away when a person of ordinary skill, upon reading the reference, would be discouraged from following the path set out in the reference, or would be led in a direction divergent from the path that was taken by the applicant." *Galderma Labs. LP v. Tolmar Inc.*, 731 F.3d 738 (Fed. Cir. 2013). However, it is not sufficient for "teaching away" if the reference "merely expresses a general preference for an alternative invention but does not criticize, discredit, or otherwise discourage investigation into the invention claimed." *Id.* The portion of Ichimura that Patent Owner relies on describes a different technique than Rosenberg for modifying the audio signal delivered to the user based on detection of the user's voice, which does not involve mixing the AC and ASM signals. However, we do not see anything in Ichimura that would discourage one of ordinary skill

from using Rosenberg's mixing of the AC and ASM signals in the context of Rosenberg's system, or in the combination of Rosenberg and Ichimura. This is particularly true because Petitioner relies on Ichimura only for detecting a cessation of voice activity and delaying modification for a predetermined time responsive to the detected cessation of voice activity in limitations 1[e] and 1[f], not the mixing of the ASM and AC signals in limitation 1[d].

Beyond the above arguments, Patent Owner does not otherwise challenge the motivation to combine Rosenberg and Ichimura. PO Resp. 9–17; PO Sur-reply 11–16. We agree with Petitioner's motivation to combine arguments, and find that one of ordinary skill would have been motivated to combine Rosenberg and Ichimura as Petitioner proposes.

Consequently, based on the foregoing, Petitioner has sufficiently proven that these limitations would have been obvious over the prior art.

g) 1[g]: "directing the mixed signal to an ear canal receiver (ECR) of the earphone device."

Petitioner argues that Rosenberg discloses or suggests this feature. Pet. 19 (citing Ex. 1002 ¶¶ 118–121). Petitioner asserts that Rosenberg discloses mixing a musical audio signal and an ambient signal into a single audio stream. Pet. 19–20 (citing Ex. 1005 ¶¶ 26, 54; Ex. 1002 ¶ 118). According to Petitioner, Rosenberg also provides that the headphones "play music directly into [the user's] ears" which one of ordinary skill "would have understood is accomplished by a speaker." *Id.* at 19–20 (citing Ex. 1005 ¶ 10; Ex. 1002 ¶ 119). "Further," Petitioner asserts, one of ordinary skill "would have understood Rosenberg's description of 'ear pieces' to disclose or suggest earphone devices in an ear canal of a user." *Id.* at 20 (citing Ex. 1002 ¶ 119). Therefore, Petitioner contends, one of ordinary skill "would have understood that Rosenberg's 'headphones (or

other similar headsets and ear pieces)' would include a speaker for providing audio to the media player user'[s] ear canal." *Id.* (citing Ex. 1002 ¶¶ 119, 120).

"Accordingly," Petitioner argues, "Rosenberg discloses or suggests outputting ('directing') a single audio stream comprising musical audio content and ambient audio content ('the mixed signal') to a media player user via speakers in ear pieces ('to an ear canal receiver (ECR) of the earphone device')." Pet. 20 (citing Ex. 1002 ¶ 121).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

h) Summary for Claim 1

Based on the full trial record, Petitioner has proven by a preponderance of the evidence that claim 1 would have been obvious based on Rosenberg in view of Ichimura.

4. Dependent Claims 2–6

Claim 2 depends from claim 1 and further recites that "the mixing of the ASM signal and the AC signal includes decreasing the ASM gain of the ASM signal and increasing the AC gain of the AC signal when the voice activity is not detected." Ex. 1001, 11:61–64. Petitioner argues that Rosenberg discloses that, after a predetermined time has elapsed during which the media player's voice is no longer detected, "the process reverses, the ambient audio content fading out to zero volume and the musical content fading back to its pre-event nominal volume." Pet. 21 (citing Ex. 1005 ¶ 58; Ex. 1002 ¶ 123).

Claim 3 depends from claim 1 and further recites that "the AC gain and the ASM gain are selected according to whether the voice activity is detected." Ex. 1001, 11:65–67. Petitioner argues that "Rosenberg discloses that, in the mixed signal, the relative volumes of the media audio signal and the ambient audio signal 'are ***selectively adjusted*** in response to detected ambient audio events,' such as, for example, 'the detection of the media player's own voice within the ambient audio signal.'" Pet. 22 (citing Ex. 1005 ¶ 55).

Claim 4 depends from claim 3 and further recites that "the mixing of the ASM and the AC signal includes: applying the ASM gain to the ASM signal to generate a modified ASM signal; applying the AC gain to the AC signal to generate a modified AC signal; and mixing the modified ASM signal and the modified AC signal to form the mixed signal." Ex. 1001, 12:1–8. Petitioner argues that "Rosenberg discloses mixing 'musical audio content' and 'ambient audio content' signals into a single audio stream, and adjusting the gains (levels) of each signal to control the relative volume in the mixed signal." Pet. 23. "In a further aspect," Petitioner asserts, "Rosenberg explains that the volume level of the musical audio content is gradually decreased to substantially zero while the volume level of the ambient audio content is gradually increased to the prior music volume level." *Id.* (citing Ex. 1005 ¶¶ 57–58). According to Petitioner, "Rosenberg further provides that 'change[s] in ***mix volumes*** may be abruptly enacted or gradually enacted,' and that the mixed audio signal with new relative volume levels lasts for a period of time, after which the volume levels of the two signals return to their pre-event nominal levels (i.e., 'the ambient audio

content fading out to zero volume and the musical content fading back to its pre-event nominal value')." *Id.* (citing Ex. 1005 ¶¶ 57–58; Ex. 1002 ¶ 127).

Claim 5 depends from claim 1, and further recites that "each of the AC gain and the ASM gain is greater than zero and less than or equal to unity gain." Ex. 1001, 12:9–11. Petitioner argues that "Rosenberg discloses mixing the media audio signal and the ambient audio signal such that 'musical audio content is gradually decreased down to substantially zero while the ambient audio musical content is gradually increased up to the prior music volume level," which is described as a "cross-fading" of the ambient and media audio signals. Pet. 25 (citing Ex. 1002 ¶ 58). According to Petitioner, one of ordinary skill "would have understood that cross-fading refers to adjusting the gain of each of the media audio signal and the ambient audio signal between zero and one such that the sum of their gains remains one." *Id.* (citing Ex. 1002 ¶¶ 134–135). Additionally, Petitioner contends, "Rosenberg discloses cross-fading such that 'the musical audio content is gradually decreased down to ***substantially zero***," and one of ordinary skill would have understood that this means that the gain of the media signal remains above zero, which necessarily means that the gain of the ambient audio signal remains below unity gain in the mixed audio signal. *Id.* at 26 (citing Ex. 1005 ¶ 58; Ex. 1002 ¶ 136).

Claim 6 depends from claim 1, and further recites that "the AC signal is received from the remote device via a wired connection or a wireless connection." Ex. 1001, 12:12–14. Petitioner argues that Rosenberg discloses the use of headphones for playing media content that "may be connected by wired or wireless connections" such as "a Bluetooth

communication link." Pet. 27 (citing Ex. 1005 ¶¶ 38, 40, 43; Ex. 1002 ¶¶ 139–140.)

Patent Owner does not present separate argument for claims 2–6. PO Resp. 25–38.

Based on the full trial record, we agree with Petitioner's arguments, and find that Petitioner has proven by a preponderance of the evidence that claims 2–6 would have been obvious over Rosenberg in view of Ichimura.

E. Ground 2A: Asserted Obviousness of Claims 7–9 Based on Rosenberg in Combination With Visser[7]

Petitioner contends that claims 7–9 would have been obvious over Rosenberg in view of Visser. Pet. 34–43. Patent Owner disagrees, arguing that Petitioner has failed to prove that claims 7–9 would have been obvious over Rosenberg and Visser. PO Resp. 38–44.

1. Overview of Visser (Ex. 1007)

Visser is titled "Systems, Methods, and Apparatus for Speech Feature Detection," and discloses "detection of a transition in a voice activity state of an audio signal, based on a change in energy that is consistent in time across a range of frequencies of the signal." Ex. 1007, codes (54), (57).

2. Independent Claim 7

Independent claim 7 is reproduced below.

> 7. [preamble] A method for passing ambient sound to an earphone device configured to be inserted in an ear canal of a user, the method comprising the steps of:

[7] Because, as discussed below, we find that Petitioner has proven by a preponderance of the evidence that claims 7–9 are unpatentable over Rosenberg in view of Visser, we need not address Petitioner's argument that claims 7–9 are unpatentable over Rosenberg in View of Lee in Ground 2B. *See Boston Scientific*, 809 F. App'x at 990.

> [a] capturing the ambient sound from an ambient sound microphone (ASM) proximate to the earphone device to form an ASM signal;
>
> [b] receiving an audio content (AC) signal from a remote device;
>
> [c] detecting voice activity of the user of the earphone device;
>
> [d] wherein the detecting of the voice activity includes: determining a time-smoothed level of a microphone signal to form a microphone level;
>
> [e] comparing the microphone level with a predetermined microphone level threshold; and
>
> [f] detecting the voice activity when the microphone level is greater than the microphone level threshold; and
>
> [g] mixing the ASM signal and the AC signal to form a mixed signal, such that, in the mixed signal, an ASM gain of the ASM signal is increased and an AC gain of the AC signal is decreased when the voice activity is detected.

Ex. 1001, 12:15–36.

Petitioner relies on its arguments for claim 1 for limitations 7[preamble], 7[a], 7[b], 7[c], and 7[g], and Patent Owner does not specifically address those limitations. Pet. 34, 41; PO Resp. 38–44. Limitations 7[d], 7[e], and 7[f] will be discussed below.

a) Limitation 7[d]

Petitioner argues that Rosenberg in combination with Visser discloses or suggests this limitation. Pet. 34 (citing Ex. 1002 ¶¶ 159–165). Petitioner asserts that Rosenberg explains that "ambient sound signals from microphone 95A are generally captured as analog audio signals and converted to digital form by an analog to digital converter or other similar component and/or process." *Id.* at 34–35 (citing Ex. 1005 ¶ 44). According to Petitioner, Rosenberg further provides that "noise reduction, filtering,

and/or other commonly known signal processing steps may be performed upon the ambient signal." *Id.* at 35 (citing Ex. 1005 ¶ 44). Once the ambient signal is converted to a final digital form, Petitioner contends, Rosenberg performs "additional signal processing" on "the captured ambient signal." *Id.* (citing Ex. 1005 ¶¶ 45–46; Ex. 1002 ¶ 159).

Petitioner argues that Visser "is directed to systems and methods for processing segments of audio signals based on whether voice activity is detected." Pet. 35 (citing Ex. 1007 ¶¶ 8–10). "For example," Petitioner asserts, "Visser describes a gain-based voice activity detection (VAD) technique 'configured to indicate presence or absence [of] voice activity in a segment based on differences between corresponding values of a gain measure for each channel,'" where "[e]ach channel is based on the signal from an array of microphones and the gain measure corresponds to the signal level picked up by a microphone." *Id.* (citing Ex. 1007 ¶¶ 113, 115, 163) (citations omitted). According to Petitioner, Visser explains that "[a] gain-based VAD technique may be configured to detect that a segment is from a desired source (e.g., to indicate detection of voice activity) when a difference between the gains of the channel is greater than a threshold value." *Id.* (citing Ex. 1007 ¶ 165). "Thus," Petitioner contends, "[b]oth Visser and Rosenberg rely on microphone signals to determine whether voice activity is present." *Id.* (citing Ex. 1002 ¶ 160).

Petitioner also points to Visser's disclosure that "[i]t may be desirable to configure the detector to perform ***a temporal smoothing operation*** on the gain measures and/or on the calculated differences." Pet. 35–36 (citing Ex. 1007 ¶ 163). Petitioner argues that these "gain measures" of the microphones "refer to levels of the microphone signals." *Id.* at 36 (citing

Ex. 1007 ¶¶ 115, 162–164; Ex. 1002 ¶ 161). According to Petitioner, one of ordinary skill would have also recognized that "[a] temporal smoothing operation" on the gain measures refers to generating a signal with a level that is a temporally-smoothed level of the gain measures, as shown in Visser's Figure 2A as an arrow from T300 to T400. *Id.* (citing Ex. 1007 ¶¶ 12, 84). Petitioner further contends that Visser shows using this temporally smoothed level in the process of voice detection. *Id.*

Petitioner argues that one of ordinary skill would have understood that "temporal" refers to "time," and that Visser's "temporally smoothed value" refers to a time-smoothed value, as supported by Visser's reference to a time derivative" as part of a method of temporal smoothing. Pet. 36 (citing Ex. 1007 ¶ 90, Fig. 2A; Ex. 1002 ¶ 162). Accordingly, Petitioner asserts, one of ordinary skill reading Visser "would have recognized that performing a 'temporal smoothing operation' on a gain measure (also called a 'level') for a channel (a microphone signal) would produce *a time-smoothed level* for the microphone corresponding to the channel, which is then used as part of the voice detection." *Id.* at 36–37 (citing Ex. 1002 ¶ 162).

Additionally, Petitioner argues that Visser discloses that using the temporally smoothed level to detect voice "help[s] to increase reliability of the onset and/or offset detection (e.g., by deemphasizing noisy artifacts).'" Pet. 37 (citing Ex. 1007 ¶ 90). For example, Petitioner asserts, "[a] non-speech sound impulse, such as a slammed door, a dropped plate, or a hand clap, may also create responses that show consistent power changes over a range of frequencies." *Id.* (citing Ex. 1007 ¶ 102). In this way, according to Petitioner, "Visser teaches that, when detecting voice activity based on a microphone signal, it is beneficial to temporarily smooth (time smooth) that

signal to improve reliability of voice detections." *Id.* Petitioner acknowledges that Visser uses more than one microphone and Rosenberg uses one microphone, but explains that it is relying on Rosenberg "for the teaching of time-smooth (and its benefit), and the rationale provided in Visser (to improve reliability of voice detection) would apply equally to the microphone 95A in Rosenberg regardless of how many microphones are described in Visser." *Id.* (citing Ex. 1002 ¶ 163).

"In view of Rosenberg's invitation to apply 'noise reduction, filtering, and/or other commonly known signal processing steps,'" Petitioner argues, one of ordinary skill "would have been motivated to modify Rosenberg's audio system based on Visser to generate a temporally-smoothed level of a microphone signal ('determine a time-smoothed level') from microphone 95A to increase the reliability of 'the onset and/or offset detection' (i.e., the detection of voice activity and cessation thereof), as Visser describes." Pet. 37–38 (citing Ex. 1002 ¶ 164). One of ordinary skill, Petitioner asserts, "would have found it obvious to combine Rosenberg and Visser in this manner because doing so would have amounted to nothing more than the application of a known technique (determining a temporally-smoothed level of a microphone signal) to a known method (Rosenberg's detection of whether a media player user or a first user is speaking) ready for improvement to yield a predictable result (using the temporally-smoothed microphone level to increase the reliability of Rosenberg's detecting whether the media player user or first user is speaking)." *Id.* at 38 (citing Ex. 1002 ¶ 164).

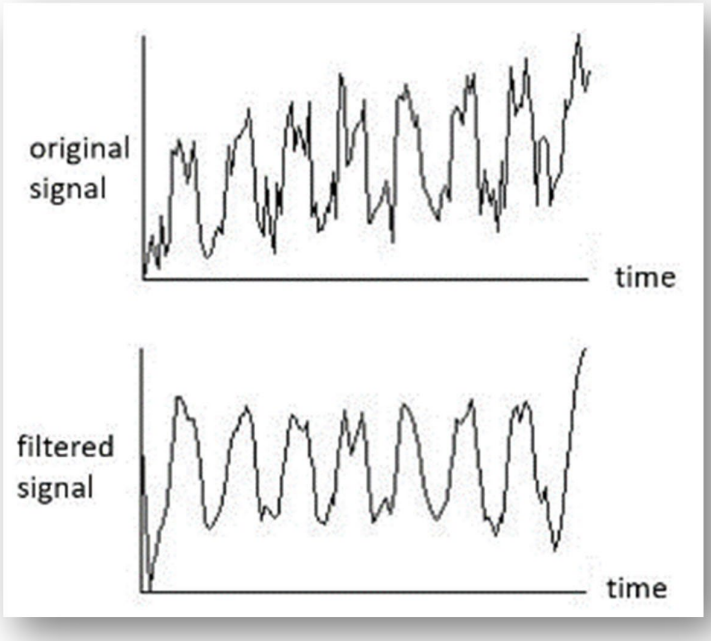
"Accordingly," Petitioner contends, "the Rosenberg-Visser system discloses or suggests wherein detecting the media player user's voice

includes ('wherein the detecting of the voice activity includes') performing a temporal smoothing operation on a level of a microphone signal ('determining a time-smoothed level of a microphone signal') to produce a time-smoothed level for a microphone ('to form a microphone level'). *Id.* (citing Ex. 1002 ¶ 165).

Patent Owner responds that one of ordinary skill would not have been motivated to modify Rosenberg based on Visser to "generate a temporally-smoothed level of a microphone signal [] from microphone 95A," because doing so "would render Rosenberg's voice detection inoperable." PO Resp. 38 (citing Pet. 37). Patent Owner argues that Rosenberg identifies characteristic forms in audio sound signals from microphone 95A using "the prior art methods of speech recognition, voice identity recognition, and environmental sound identification." *Id.* (citing Ex. 1005 ¶¶ 39, 46). Patent Owner asserts that, for the characteristic form relating to a user's voice, Rosenberg discloses that the identification is "performed upon a certain time-sample's worth of ambient audio signal," and "[t]he experts agree that [a person of ordinary skill in the art would have understood this as a description of amplitude as a function of time, and would apply this description to all three characteristic forms." *Id.* at 38–39 (citing Ex. 1005 ¶ 46; Ex. 2007, 31:17–32:18).

Patent Owner acknowledges that "Rosenberg contemplates some signal processing techniques on the ambient signal such as 'noise reduction, filtering, and/or other commonly known signal processing steps,'" but argues that "temporal smoothing is not one of them." Pet. 40 (citing Ex. 1005 ¶ 44). Indeed, Patent Owner argues, "Rosenberg relies on the uncorrupted microphone signal to detect the 'unique voice of the media

player user,' which would require spectral analysis," and one of ordinary skill would have understood that temporal smoothing of Rosenberg's microphone signal would "artificially change[]" the spectrum, which would "render the signal unusable in the spectral analysis." *Id.* (citing Ex. 2006 ¶ 110). Patent Owner provides a figure to illustrate the difference between an example signal's spectral profile showing amplitude as a function of time before and after temporal smoothing, which is reproduced below.



Patent Owner's figure showing the difference between an example spectral profile before and after temporal smoothing. PO Resp. 41 (citing Ex. 2006 ¶ 111).

Patent Owner argues that, as shown in the above figure, a smoothed signal would have a different spectral profile than the original signal, including fewer high-frequency components and more low-frequency components. *Id.* at 41 (citing Ex. 2006 ¶ 112). Thus, Patent Owner contends, "temporal smoothing likely would render Rosenberg unable to identify the unique

voice of the media player user." *Id.* at 41–42 (citing Ex. 2006 ¶ 112; Ex. 2007, 49:4–49:18).

Patent Owner further argues that the proposed Rosenberg-Visser system would attempt to identify the user's voice by comparing the smoothed sound profile with Rosenberg's existing, unsmoothed sound profiles, which would "at a minimum, yield errors including false detection and failure to detect user voices." Pet. 42 (citing Ex. 2006 ¶ 113). From this, Patent Owner asserts, one of ordinary skill would have known to avoid smoothing Rosenberg's signal because "otherwise, Rosenberg would lose the ability to detect the user's voice," rendering Rosenberg "inoperative." *Id.* (citing Ex. 2006 ¶ 115).

Petitioner responds that Patent Owner's assertion that Rosenberg requires spectral analysis is incorrect because Rosenberg "incorporates by reference example prior art techniques to perform speech recognition, voice identity recognition and environmental sound identification, and states that these prior art techniques can be used in its system." Pet. Reply 17 (citing PO Resp. 40; Ex. 1005 ¶¶ 45–46). For example, Petitioner argues, Rosenberg identifies and incorporates by reference U.S. Patent No. 4,054,749 to Suzuki, which uses time domain-based techniques for voice recognition. *Id.* at 18 (citing Ex. 1005 ¶¶ 45; Ex. 1028 ¶¶ 17–21). Thus, according to Petitioner, "Rosenberg explicitly discloses using time domain-based techniques for performing user voice recognition." *Id.* Petitioner further contends that Mr. Kleinschmidt admitted at his deposition that he had not reviewed any of the signal analysis reference that Rosenberg incorporates by reference. *Id.* (citing Ex. 1025, 90:13–91:18).

Petitioner also argues that even if one were to assume that Rosenberg must use spectral analysis for user voice identity recognition, using temporal smoothing would not render the signal unusable in the spectral analysis. Pet. Reply 18–19 (citing PO Resp. 40, 44). Petitioner asserts that Rosenberg discloses performing "filtering" on the ambient signal before performing voice identity reduction, and that filtering can extract extraneous noise that is outside certain frequency limits. *Id.* at 19 (citing Ex. 1005 ¶¶44–46, 54; Ex. 1028 ¶ 24). Petitioner also argues that Mr. Kleinschmidt admitted that he "knows nothing about" the signals in Patent Owner's figure showing the spectral profile of an exemplary signal's amplitude as a function of time, and that these signals "were simply provided to him by counsel." *Id.* at 20–21 (citing PO Resp. 41, 43; Ex. 2006 ¶¶ 111, 116; Ex. 1025, 92:16–94:8). According to Patent Owner, one of ordinary skill would have no basis to conclude that this signal "had been smoothed to such an extent so as to be incapable of being used for voice identity recognition" and, in any event, Rosenberg "contemplates altering the ambient microphone signal's spectrum before performing voice identity recognition." *Id.* at 21–22 (citing Ex. 1028 ¶ 28).

Finally, Petitioner argues that "Rosenberg refutes [Patent Owner's] argument that 'smoothing the signal received by Rosenberg should be avoided; otherwise, Rosenberg would lose the ability to detect the user's voice." Pet. Reply 23 (citing PO Resp. 42; Ex. 1028 ¶¶ 32–33). According to Petitioner, Rosenberg discloses filtering the ambient audio signal (which smooths it), and one of ordinary skill would have know or found it obvious to filter microphone signals used to create sound profiles (to the extent sound profiles were required) to enable a comparison of the ambient

microphone signal to the user profile. *Id.* at 22–23 (citing Ex. 1005 ¶¶ 46, 54; Ex. 1028 ¶¶ 31–34).

Patent Owner responds that Petitioner's arguments based on the Suzuki patent were not mentioned in the Petition or Dr. Polish's original declaration, and should not be considered. PO Sur-reply 16 (citing Ex. 1002 ¶¶ 6–7). Patent Owner also argues that spectral analysis was a common technique for voice identification, and even if it were possible to perform voice identification without it, avoiding spectral analysis "would have frustrated Rosenberg for the vast majority of implementations." *Id.* at 17–18. Additionally, Patent Owner asserts that Petitioner's criticisms of Mr. Kleinschmidt's lack of knowledge about the time smoothed signals in Petitioner's diagram is "trivial" because the precise algorithm used to generate the time smoothed signal is "irrelevant." *Id.* at 18. Finally, Patent Owner contends that Petitioner pieces together disparate disclosures of "filtering" in Rosenberg, and fails to show that these disclosures related to voice identification. *Id.* (citing Ex. 1005 ¶¶ 44, 54; Pet. Reply 19).

Based on the full trial record, we find that Petitioner has sufficiently proven that this limitation would have been obvious based on the combination of Rosenberg and Visser. Rosenberg discloses that "noise reduction, filtering, and/or other commonly known signal processing steps may be performed upon the ambient signal," and that "the ambient audio signal content may be filtered or otherwise processed to extract extraneous noise" that "is outside certain magnitude and/or frequency limits or threshold." Ex. 1005 ¶¶ 44, 54. Rosenberg discloses that the system them performs "additional signal processing" on "the captured ambient signal" which "may include sound recognition processing, speech recognition

processing, and/or vocal identity recognition processing steps and/or substeps." *Id.* ¶ 45. Rosenberg further explains that, "speech recognition, and/or vocal identity recognition processes are known to the prior art," and incorporates by reference a number of patents and references describing such processes. *Id.*

Like Rosenberg, Visser is directed to systems and methods for processing segments of audio signals based on whether voice activity is detected by a microphone. Ex. 1007 ¶¶ 8–10; Ex. 1002 ¶ 160. Visser also discloses that "[i]t may be desirable to configure the detector to perform a ***temporal smoothing operation*** on the gain measures and/or on the calculated differences." Ex. 1007 ¶ 163 (emphasis added). We agree with and find credible Dr. Polish's testimony that "[t]he 'gain measures' of the microphones discussed in Visser refer to levels of the microphone signals," and thus one of ordinary skill "would recognize that '[a] temporal smoothing operation' on the gain measures refers to generating a signal with a level that is a temporally-smoothed level of the gain measures." Ex. 1002 ¶ 161 (citing Ex. 1007 ¶¶ 162–164). Additionally, Visser shows the use of this temporally smoothed level in the process of voice detection. Ex. 1007 ¶¶ 12, 84; Ex. 1002 ¶ 161. Thus, we agree with Dr. Polish that Visser discloses a time-smoothed level for the microphone corresponding to a channel, which may then be used as part of voice detection. Ex. 1002 ¶ 162.

Visser further discloses that using a temporally-smoothed level to detect voice "help[s] to increase reliability of the onset and/or offset detection (e.g., by deemphasizing noisy artifacts)." Ex. 1007 ¶ 90; Ex. 1002 ¶ 163. For example, Visser explains that "[a] non-speech sound impulse, such as a slammed door, a dropped plate, or a hand clap, may also create

responses that show consistent power changes over a range of frequencies." Ex. 1007 ¶ 102; Ex. 1002 ¶ 163. We agree with and find credible Dr. Polish's testimony that, based on these disclosures, one of ordinary skill would have understood that "Visser teaches that, when detecting voice activity based on a microphone signal, it is beneficial to temporally smooth (time smooth) the signal to improve reliability of voice detections." Ex. 1002 ¶ 163. We also agree with Dr. Polish that, "[a]lthough Visser describes more than one microphone, and Rosenberg describes a single microphone," the "rationale provided in Visser (to improve reliability of voice detection) would apply equally to the microphone 95A in Rosenberg." *Id.* Additionally, we credit Dr. Polish's testimony that "in view of Rosenberg's invitation to apply 'noise reduction, filtering, and/or other commonly known signal processing steps,'" one of ordinary skill "would have been motivated to modify Rosenberg's audio system based on Visser to generate a temporally smoothed level of a microphone signal ('determine a time smoothed level') from microphone 95A to increase the reliability of 'the onset and/or offset detection' (*i.e.*, the detection of voice activity and cessation thereof), as Visser describes." *Id.* ¶ 164.

We do not agree with Patent Owner's argument that modifying Rosenberg based on Visser to generate a temporally-smoothed microphone signal would render Rosenberg's voice detection inoperable. *See* PO Resp. 38. Although we agree with Mr. Kleinschmidt that temporal smoothing of the microphone signal would change the signal profile (at least to some degree), he fails to provide a sufficient explanation of why the Rosenberg-Visser system would be unable to use this smoothed profile for voice recognition. Ex. 2006 ¶¶ 110–114. Mr. Kleinschmidt states that the

temporally smoothed signal would have higher frequencies reduced and lower frequencies enhanced (*id.* ¶ 110), but that does not mean that the signal would be unrecognizable. Mr. Kleinschmidt further asserts that comparing smoothed sound profiles to unsmoothed profiles could result in errors (*Id.* ¶ 113), but that also does not mean that the system would be inoperable. As for Dr. Kleinschmidt's contention that in the proposed combination the smoothed profile would have to be compared to Rosenberg's existing, unsmoothed profile, we find more credible the testimony of Dr. Polish that one of ordinary skill "would have known or at a minimum found it obvious to similarly filter the microphone signal used to create the user's sound profile(s) in order to meaningfully enable a comparison of the ambient microphone signal to the user profile." Ex. 1002 ¶ 113; Ex. 1028 ¶ 34.

Additionally, we find that Mr. Kleinschmidt's opinion that temporal smoothing would destroy Rosenberg's ability to recognize a voice signal is contradicted by the testimony of Dr. Polish, which we find more credible and persuasive. *See* Ex. 1028 ¶¶ 31–33. Specifically, we are persuaded by Dr. Polish's testimony that "Rosenberg discloses filtering the ambient audio signal, and a person of ordinary skill in the art would have known that low-pass filtering the signal time smooths it," indicating that low-pass filtering the audio signal "need not be avoided, and indeed may be desirable (e.g., to remove noise)." *Id.* ¶ 31. Moreover, as Dr. Polish explains, "Rosenberg explicitly discloses performing 'noise reduction, filtering, and/or other commonly known signal processing steps' on the ambient microphone signal before processing the signal to perform voice identity recognition," indicating that Rosenberg contemplates signal processing operations that

alter the ambient microphone signal prior to performing voice identity recognition. *Id.* ¶ 32.

We further disagree with Patent Owner's argument that Petitioner pieces together disparate disclosures of filtering in Rosenberg and fails to show that these disclosures relate to voice identification. *See* PO Sur-reply 17–18. Petitioner and Dr. Polish cite to paragraphs 44, 45, and 54 of Rosenberg which, as discussed above, describe filtering the ambient audio signal to extract noise, and then performing "additional signal processing" on "the captured ambient signal" which "may include sound recognition processing, speech recognition processing, and/or vocal identity recognition processing steps and/or substeps." Ex. 1005 ¶¶ 44, 45, 54; Pet. 34–35; Ex. 1028 ¶ 32.

Based on the foregoing, Petitioner has sufficiently proven that this limitation would have been obvious over the prior art.

b) Limitation 7[e]

Petitioner argues that Rosenberg in combination with Visser discloses or suggests this limitation. Pet. 38 (citing Ex. 1002 ¶¶ 166–170). For example, Petitioner explains, Rosenberg describes enabling a user to configure a media player "to be responsive only to name utterances that exceed a certain volume threshold" and, in order to do so, one of ordinary skill "would have recognized that the media player would need to compare the volume of the ambient audio signal to a certain volume threshold ('a predetermined microphone level threshold')." *Id.* at 38–39 (citing Ex. 1005 ¶¶ 53; Ex. 1002 ¶ 166).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 38–44.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

c) Limitation 7[f]

Petitioner argues that Rosenberg in combination with Visser discloses or suggests this limitation. Pet. 40 (citing Ex. 1002 ¶ 171). Petitioner asserts that "Rosenberg discloses or suggests detecting a media player user's own voice ('voice activity of the user of the earphone device') within a captured ambient audio signal." *Id.* Petitioner asserts that "Rosenberg provides that 'the ambient audio signal content may be filtered or otherwise processed to extract extraneous noise and/or sound content that is outside certain magnitude and/or frequency limits *or thresholds*.'" *Id.* (citing Ex. 1005 ¶ 54). Additionally, according to Petitioner, "[c]laim 8 of Rosenberg describes that the device is responsive if 'a volume of the one or more characteristic forms exceeds a volume threshold.'" *Id.* at 40–41 (citing Ex. 1005 ¶ 54). For reasons previously discussed with respect to limitation 7[e], Petitioner argues, "the Rosenberg-Visser system discloses or suggests detecting the media user player's voice ('detecting the voice activity') when the temporally-smoothed signal from microphone 95A ('when the microphone level') is above a predetermined volume level threshold ('is greater than the microphone level threshold')." *Id.* at 41 (citing Ex. 1002 ¶ 171).

Patent Owner does not present arguments directed to this limitation. *See* PO Resp. 38–44.

Based on the full trial record, we agree with Petitioner's arguments, and find that this limitation would have been obvious over the prior art.

d) Summary for Claim 7

Based on the full trial record, Petitioner has proven by a preponderance of the evidence that claim 7 would have been obvious based on Rosenberg in view of Visser.

3. Dependent Claims 8 and 9

Claim 8 depends from claim 7 and further recites:

> The method according to claim 7, wherein the detecting of the voice activity includes detecting the voice activity from the microphone signal, the microphone signal including at least one of the ASM signal or an ear canal microphone (ECM) signal captured within the ear canal from an ECM of the earphone device.

Ex. 1001, 12:37–43.

Petitioner argues that Rosenberg in combination with Visser discloses or suggests the features of claim 8. Pet. 41 (citing Ex. 1002 ¶¶ 173–175). Petitioner asserts that "Rosenberg discloses a microphone 95A to detect ambient sounds," and "provides for the identification of 'the unique voice of the media player user' or the voice of another user by analyzing the signal from microphone 95A." *Id.* (citing Ex. 1005 ¶ 39, 46; Ex. 1002 ¶ 173). "Accordingly," Petitioner contends, "the Rosenberg-Visser system discloses or suggests detecting a media player user's own voice ('voice activity') within the ambient audio signal from microphone 95A ('from the microphone signal including at least one of the ASM signal or an ear canal microphone . . .')." *Id.* at 41–42 (citing Ex. 1002 ¶¶ 174–175).

Dependent claim 9 depends from claim 8, and further recites "filtering at least one of the microphone signal or the AC signal by a predetermined filtering characteristic." Ex. 1001, 12:43–45. Petitioner argues that Rosenberg discloses that "noise reduction, ***filtering***, and/or other commonly

known signal processing steps may be performed upon the ambient signal." Pet. 42 (citing Ex. 1005 ¶ 39). And, Petitioner explains, "[o]nce converted to a final digital form, Rosenberg explains that 'additional signal processing is performed on the captured ambient signal.'" *Id.* (citing Ex. 1005 ¶ 45). According to Petitioner, one of ordinary skill "would have recognized that 'filtering . . . performed upon the ambient signal' would comprise filtering based on a predetermined filtering characteristic." *Id.* (citing Ex. 1002 ¶ 177).

Patent Owner does not present separate argument for claims 8–9. PO Resp. 38–44.

Based on the full trial record, we agree with Petitioner's arguments, and find that the limitations of claims 8–9 would have been obvious over Rosenberg in view of Visser.

F. Ground 3: Asserted Obviousness of Claim 10 Based on Rosenberg, Visser, and Kvaløy

Petitioner contends that claim 10 would have been obvious over Rosenberg in view of Visser and Kvaløy. Pet. 47–51. Patent Owner disagrees. PO Resp. 51–67.

1. Overview of Kvaløy (Ex. 1009)

Kvaløy is directed to "a voice detection and discrimination apparatus in a hearing protection arrangement." Ex. 1009, 1:16–18. The apparatus "is intended for use in noisy environments," such as those near heavy operating machinery, loud vehicle traffic, or crowds of people. *Id.* at 1:30–39. An embodiment of Kvaløy's apparatus included in an earplug is shown in Figure 1, reproduced below.

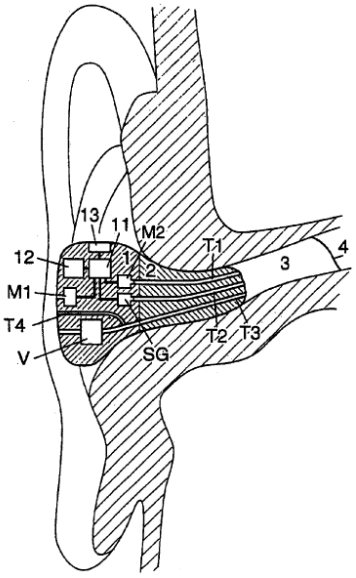


Fig. 1

Figure 1 shows an example of Kvaløy's voice detection and discrimination apparatus. Ex. 1009, Fig. 1, 3:38–39.

As shown in Figure 1, Kvaløy's voice detection and discrimination apparatus includes two electroacoustic transducer elements M1 and M2. Ex. 1009, 3:41–43. The sound inlet of M1 is connected to the outside of the earplug and picks up external sounds, while M2 is connected to the inner portion of the meatus 3 by means of an acoustic transmission channel T1. *Id.* at 3:52–57. The apparatus also includes sound generator SG that "is open into the inner portion of the meatus 3 by means of an acoustic transmission channel T2 between the sound generator SG and the inward facing portion of sealing section 2." *Id.* at 3:62–66. Additionally, the apparatus includes a "sealing part" that is "made of a resilient, slowly re-expanding shape retaining polymer foam like PVC, PUR, or other materials suitable for earplugs." *Id.* at 4:21–26.

A block diagram of the main functional units of the electronic circuitry of Kvaløy's apparatus is shown in Figure 2, reproduced below.

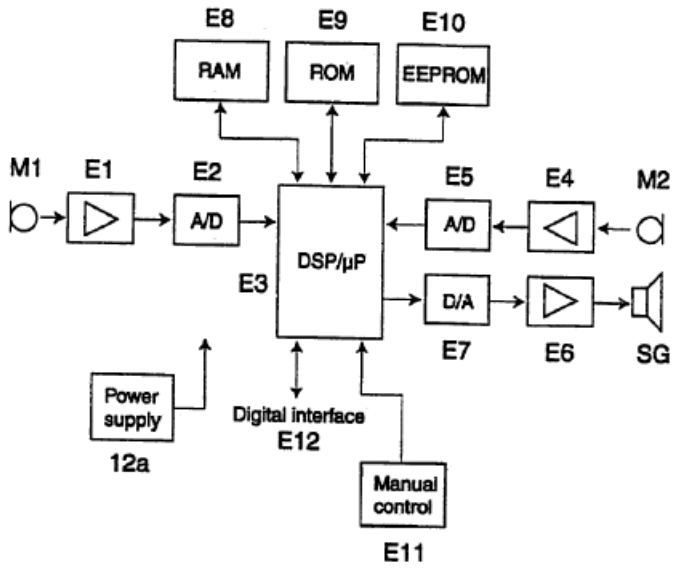


Fig. 2

Figure 2 shows a block diagram of the main functional units of the electronic circuitry of Kvaløy's apparatus. Ex. 1009, Fig. 2, 5:1–3.

As shown in Figure 2, Kvaløy's apparatus includes "signal processing unit E3," "storage means which may be RAM (Random access memory) E8, ROM (read only memory) E9, or EEPROM (electrically erasable programmable read only memory) E10, or combinations of these." Ex. 1009, 5:10–11, 5:50–57. The apparatus also includes "bi-directional digital interface E12," which may use "[t]he Bluetooth standard" for wireless communication and can allow the "two ear terminals 1,2" to be "used in a binaural mode." *Id.* at 5:66–6:18. Signal processor E3 may generate an output signal for sound generator SG based on "signals received by the electronic circuitry 11 via the communication with other electrical units" or "signals detected by the electroacoustic transducer elements M1, M2." *Id.* at 6:21–24. Signal processing unit E3 can also include "signal analysis means for detecting the presence of speech components, such as

words, in the signal from the inner electroacoustic transducer element M2." *Id.* at 9:29–36; *see id.* at 6:36–41.

2. Dependent Claim 10

Claim 10 depends from claim 7, and further recites:

> The method according to claim 7, wherein the detecting of the voice activity includes: determining a time-smoothed level of the AC signal to form an AC level; comparing the AC level with an AC level threshold; and detecting the voice activity when the microphone level is greater than the microphone level threshold and the AC level is less than the AC threshold.

Ex. 1001, 12:46–53.

Petitioner argues that *Rosenberg* in combination with Visser and Kvaløy discloses or suggests the features of claim 10. Pet. 47 (citing Ex. 1002 ¶¶ 197-206). Petitioner notes that these features are "similar to the method recited in claim 7, in which the detection of voice activity includes time-smoothing the ambient audio signal and comparing the time-smoothed signal to a threshold." *Id.* Claim 10, Petitioner asserts, further requires "performing this process on the audio content (media content) signal, in addition to the ambient sound signal, and only detecting voice when the time-smoothed audio signal is below a threshold, rather than above it." *Id.* (citing Ex. 1002 ¶ 197).

Petitioner argues that "Kvaløy describes this feature in the context of a 'protective ear terminal element' designed for 'situations where it is desirable for people to use a hearing protection arrangement, while still requiring some means of communicating, e.g., to speak with other people.'" Pet. 47 (citing Ex. 1009, 1:39–44, 2:39–40; Ex. 1002 ¶ 198) (citations omitted). Petitioner asserts that "Kvaløy provides for 'voice detection' in noisy environments 'with improved voice detection capability and which has

a reduced false activation due to acoustic noise.'" *Id.* (citing Ex. 1009, 2:14–24; Ex. 1009 ¶ 198). According to Petitioner, Kvaløy provides an embodiment "where the voice detection and discrimination is included in an earplug based hearing protective voice communication terminal." *Id.* at 47–48 (citing Ex. 1009, 3:39–41).

Petitioner argues that Kvaløy's earplugs "receive a 'communication signal' (audio content) and produce a corresponding sound signal to the eardrum via a loudspeaker." Pet. 48 (citing Ex. 1009 5:25–32; Ex. 1002 ¶ 199)." Petitioner asserts that Kvaløy "notes that when the incoming communication signal is introduced (audio content is played) in the same terminal (earphone) as is used for voice activation control (detection of voice activity), 'it is necessary to apply a blocking function' to the voice detection process that 'depends on the incoming communication signal.'" *Id.* (citing Ex. 1009 5:33–38). According to this blocking function, Petitioner asserts the "additional decision condition signal" will "prohibit or block the detection of the incoming communication signal as if it were the users own voice, during the periods of time when the incoming communication signal is active.'" *Id.* (citing Ex. 1009, 5:38–42; Ex. 1002 ¶ 200). Thus, Petitioner contends, "Kvaløy describes blocking detection of voice when the audio content is being played, so that the system does not confuse the communication signal (audio content) for the user's voice." *Id.* (citing Ex. 1009 2:21–24; Ex. 1002 ¶ 201).

Petitioner argues that one of ordinary skill "would have readily recognized that the Rosenberg-Visser system would be improved with the blocking feature in Kvaløy" in which "voice detection is blocked when audio content being played from the device is loud enough to be confused by

the system 'as if it were the user's own voice.'" Pet. 48 (citing Ex. 1009, 5:38–42; Ex. 1002 ¶¶ 202–204). Petitioner asserts that "Rosenberg describes an automatic ambient pass-through feature that increases the volume of ambient audio signal and reduces the media audio signal (e.g., music) when, e.g., the user's voice is detected." *Id.* at 49. Kvaløy, according to Petitioner, "teaches that an incoming signal played from a loudspeaker in a voice-detecting ear-piece can cause 'false activation' of the voice-detection feature." *Id.* (citing Ex. 1009, 2:21–24; Ex. 1002 ¶ 205). One of ordinary skill, Petitioner contends, "would have found it desirable to avoid these false activations in the Rosenberg-Visser system due to audio content being played out of the ear pieces and mistakenly being detected 'as if it were the user's own voice,'" and "Kvaløy teaches that this can be accomplished by blocking the voice detection feature when the audio content is being played above a certain threshold." *Id.* (citing Ex. 1009, 5:38–42; Ex. 1002 ¶ 205). Accordingly, Petitioner asserts, "in addition to detecting voice activity when the ambient microphone signal level is above a threshold," one of ordinary skill "would also have recognized the added benefit of the blocking feature in Kvaløy, i.e., detecting voice activity when the media audio signal is below a threshold." *Id.* at 49–50 (citing Ex. 1002 ¶ 205). Additionally, Petitioner contends, "[i]n implementing this blocking feature," one of ordinary skill "would have been motivated to use a time-smoothed level of the media audio signal as taught by Visser." Pet. 50 (citing Ex. 1002 ¶ 206).

Patent Owner responds that Kvaløy's blocking feature is incompatible with Rosenberg. PO Resp. 51. According to Patent Owner, "Petitioner['s] obviousness argument against claim 10 hinges on the condition that

Kvaløy's blocking function comes with a threshold" that "Kvaløy neither discloses nor suggests." *Id.* at 52. To the contrary, Petitioner argues, Kvaløy's blocking occurs "during the periods of time when the incoming signal is active," and thus "Kvaløy implements blocking anytime an incoming signal exists." *Id.* (citing Ex. 1009, 5:38–42). Thus, Patent Owner asserts, the Rosenberg-Visser-Kvaløy combination would "block 'the voice detection feature' of Rosenberg any time the system receives media audio signals (i.e., whenever the user plays music)." *Id.* As a result, according to Petitioner, the combined system "would never detect when the user speaks, and the volume changes that Rosenberg desires when the user speaks would not be implemented." *Id.* at 53 (citing Ex. 1005 ¶ 48; Ex. 2006 ¶ 135).

Patent Owner also argues that Petitioner improperly adds to Kvaløy a disclosure that the blocking function is not a blanket prohibition, but rather is conditional. PO Resp. 53–67. Patent Owner asserts that Dr. Polish suggests that voice recognition could be used to determine whether to activate the blocking function, but that this is a new argument, is unsupported by Kvaløy's disclosure, and would add significant complexity. PO Resp. 53–64. According to Patent Owner, Kvaløy's blocking feature "has no use for voice recognition" because it looks for "a difference between two signal strengths" and interprets "a decrease in signal difference" as "a voice signal being present." *Id.* at 62 (citing Ex. 1009, 4:52–53). Patent Owner also asserts that this portion of Kvaløy discusses "voice activity detection," which merely detects voice activity, rather than "vocal signal identification/recognition," which "compare[s] a sound signal to a stored value using [a] spectrum to match, for example, keywords." *Id.* at 63 (citing Ex. 2006 ¶ 146).

Petitioner responds that Patent Owner's argument that Kvaløy's blocking function "has no threshold or any other condition" and would block Rosenberg's "voice detection feature" any time the system receives media audio signals" mischaracterizes Kvaløy's disclosure. Pet. Reply 24–25. Petitioner argues that Kvaløy discloses "an additional decision condition signal applied to decision block 28 to ensure that the system continues to function properly to recognize user voice inputs even in the presence of an incoming communication signal, ***not*** a blocking function to destroy the system's ability to recognize user voice inputs whenever an incoming communication signal is present." *Id.* at 26. Petitioner points to Kvaløy's disclosure that the blocking function is intended to "block the detection of the incoming communication signal ***as if it were the user's own voice***," and argues that Kvaløy does not disclose or suggest that "the blocking function is designed to block the user from being able to provide user voice commands when an incoming communication signal is active." *Id.* (citing Ex. 1009, 5:35–42).

Petitioner further argues that "Kvaløy states that the blocking function is for an ear terminal 'used for voice activated control' and that the 'additional decision control signal' '***typically depends on*** the incoming communication signal.'" Pet. Reply 27 (citing Ex. 1009, 5:33–44). According to Petitioner, "[i]f Kvaløy meant to disclose a binary on/off decision," the "'decision control signal' would not 'typically depend' on the incoming communication signal." *Id.* "Furthermore," Petitioner asserts, "Kvaløy discloses that the signal from the inner microphone does not need to be completely free of any noise/feedback component from the speaker because the 'processing unit E3' includes 'signal analysis means' that

'comprises means for ***separating the voice signal from the total signal*** detected by the inner electroacoustic transducer element.'" *Id.* at 27–28 (citing Ex. 1009, 9:37–43).

Patent Owner responds that the parties' dispute for this ground "boils down to a single question: How does Kvaløy's blocking function 'prohibit or block the detection of the incoming communication signal as if it were the user[']s own voice, during the periods of time when the incoming communication signal is active'?" PO Sur-reply 19 (citing Ex. 1009, 5:35–42). According to Patent Owner, Petitioner improperly "assume[s] that Kvaløy employs a sophisticated user voice recognition scheme that distinguishes between an incoming communication signal (such as music in the proposed Rosenberg-Visser-Kvaløy combination) and the user's own voice," but "Kvaløy says otherwise." *Id.* at 20 (citing Pet. Reply 27–28). Patent Owner contends that Kvaløy's "signal analysis means" in "processing unit E3" does not apply to decision block 28, which makes a decision based on the signal strength difference between M1 and M2 that is "***independent of the sound character and sound level***." *Id.* (citing Ex. 1009, 8:13–17. Thus, Patent Owner asserts, Kvaløy's blocking function does not block the microphone signal itself because Kvaløy's system cannot distinguish the user's voice from another within the same microphone signal." *Id.*

Based on the full trial record, we find that Petitioner has failed to meet its burden to prove that the Rosenberg-Visser-Kvaløy combination teaches claim 10. To begin with, Dr. Polish's testimony appears somewhat unclear as to the operation of Kvaløy's blocking function. Dr. Polish first testifies that one of ordinary skill "would recognize that Kvaløy describes blocking detection of voice when the audio content is being played so that the system

does not confuse the communication signal (audio content) for the user's voice." Ex. 1002 ¶ 200. Later in his declaration, however, Dr. Polish opines that one of ordinary skill would have recognized that the Rosenberg-Visser system "would be improved with the blocking feature in Kvaløy whereby voice detection is blocked ***when audio content being played from the device is loud enough to be confused by the system*** 'as if it were the user's own voice.'" *Id.* ¶ 205 (citing Ex. 1009, 5:38–42) (emphasis added); *see also* Pet. 48. The first statement appears consistent with Patent Owner's understanding that the blocking function is binary and not conditional, while the second interprets Kvaløy as disclosing a conditional blocking feature that only blocks voice detection when the audio content being played is sufficiently loud.

To the extent Petitioner and Dr. Polish are interpreting Kvaløy's blocking function as being conditional based on the loudness of the incoming communication signal, that interpretation is not supported by Kvaløy's disclosure. The cited portion of Kvaløy states as follows:

> When the incoming communication signal is introduced in the same terminal as used for voice activated control, it is necessary to apply a blocking function in the form of an additional decision condition signal to the decision process. This additional decision condition signal typically depends on the incoming communication signal. The additional decision condition signal will prohibit or block detection of the incoming communication signal as if it were the user[']s own voice, during the periods of time when the incoming communication signal is active.

Ex. 1009, 5:32–42. We see nothing in this portion of Kvaløy that discloses or suggests that the blocking feature only blocks detection of the user's voice when the audio content being played from the device is of a sufficient loudness. The above-quoted portion does say that the decision condition

signal will block detection of the incoming communication signal "as if it were the user[']s own voice," but this does not disclose or suggest that the blocking function depends on the loudness of the incoming control signal. Similarly, Petitioner relies on Kvaløy's statement that the additional decision control signal "typically depends on the incoming communication signal," but that also does not suggest a loudness condition for the blocking function.

Our conclusion in this regard is supported by the testimony of Mr. Kleinschmidt, which we find credible. Specifically, Mr. Kleinschmidt testifies that the above portion of Kvaløy discloses that "Kvaløy implements blocking where any amount of incoming communication signal is active, regardless of its volume." Ex. 2006 ¶ 133; *see also id.* ¶ 132 (testifying that, "[i]n my opinion, Kvaløy does not teach any determination based on a threshold; rather, Kvaløy's blocking feature is always on any time an incoming communication signal is being received"). Mr. Kleinschmidt further explains that "Kvaløy's system does not, and does not need to, analyze whether an incoming communication signal meets other conditions to activate the blocking function." *Id.* ¶ 148. Moreover, to the extent Petitioner is arguing that one of ordinary skill would have been motivated to modify Kvaløy to add a loudness condition to the blocking function, we find that such a modification is not supported by the references or the testimony of Dr. Polish. Furthermore, we find that Petitioner's argument that Kvaløy's additional decision condition signal applied to decision block 28 "ensure[s] that the system continues to function properly to recognize user voice inputs even in the presence of an incoming communication signal" is unsupported by Kvaløy, Dr. Polish, or other evidence in the record. *See* Pet. Reply 26.

We also are not persuaded by Petitioner's argument that "Kvaløy discloses that the signal from the inner microphone does not need to be completely free of any noise/feedback component from the speaker because the 'processing unit E3' includes 'signal analysis means' that 'comprises means for separating the voice signal from the total signal detected by the inner electroacoustic transducer element.'" Pet. Reply 27–28 (citing Ex. 1009, 9:37–43) (emphasis omitted). The cited portion of Kvaløy does not discuss the blocking function, and does not suggest that the ability to detect words or separate out the voice signal obviates the use of Kvaløy's blocking function which blocks the signal from the inner microphone when the incoming communication signal is active. *See* Ex. 1009, 9:29–43. Moreover, to the extent Petitioner argues that Kvaløy's blocking function is conditioned upon voice recognition, Mr. Kleinschmidt persuasively explains that Kvaløy's voice detection in Figure 3 only involves a comparison of signal strengths from Kvaløy's inner and outer microphones, and does not use speech or voice recognition. Ex. 2006 ¶¶ 138–143; Ex. 1009, 4:42–55, Fig. 3.

Furthermore, we agree with and find credible Mr. Kleinschmidt's testimony that combining Kvaløy's blocking function into the proposed combination would interfere with Rosenberg's operation because it would "block 'the voice detection feature' of Rosenberg any time the system receives media audio signals (i.e., whenever the user plays music)." Ex. 2006 ¶ 134. In other words, "if a user of the proposed Rosenberg-Visser-Kvaløy combination is listening to music, voice detection will be deactivated (so that the system does not mistake the music as the user's voice)." *Id.* Thus, as Mr. Kleinschmidt persuasively explains, "when the

user of the proposed Rosenberg-Visser-Kvaløy combination is listening to music, the system would never detect when the user speaks, and the volume changes that Rosenberg desires when the user speaks would not be implemented." *Id.* ¶ 135. Similarly, Kvaløy's blocking function would appear to interfere with the combination's ability to "detect the other characteristic forms disclosed in Rosenberg," such as "another person uttering the user's name and an emergency related alert sound" when "the user is listening to music." *Id.* ¶ 136.

Consequently, Petitioner has failed to prove by a preponderance of the evidence that one of ordinary skill would have been motivated to combine Kvaløy's blocking function with Rosenberg and Visser, or that claim 10 would have been obvious over the combination of Rosenberg, Visser, and Kvaløy.

G. Ground 4B: Asserted Obviousness of Claims 13–20 Based on Rosenberg, Kvaløy, Visser, and Ichimura[8]

1. Independent Claim 13

Independent claim 13 is reproduced below.

1. [preamble] An earphone system comprising:

[a] at least one earphone device including:

[b] a sealing section configured to conform to an ear canal of a user of the earphone device;

[c] an ear canal receiver (ECR);

[8] Because, as discussed further below, we find that Petitioner has proven by a preponderance of the evidence that claims 13–18 are unpatentable over the combination of Rosenberg, Kvaløy, Visser, and Ichimura in Ground 4B, we need not decide whether Petitioner has also proven that claims 13–18 are unpatentable over Rosenberg, Kvaløy, and Visser in Ground 4A. *See Boston Scientific*, 809 F. App'x at 990.

[d] an ambient sound microphone (ASM) for capturing ambient sound proximate to the earphone device and to form an ASM signal;

[e] a signal processing system configured to:

[f] receive an audio content (AC) signal from a remote device,

[g] detect voice activity of the user of the earphone device,

[h] mix the ASM signal and the AC signal to form a mixed signal, such that, in the mixed signal, an ASM gain of the ASM signal is increased and an AC gain of the AC signal is decreased when the voice activity is detected, and

[i] direct the mixed signal to the ECR; and

[j] a voice activity detector (VAD) timer system configured to:

[k] detect a cessation of the voice activity, and

[l] delay modification of the ASM gain and the AC gain for a predetermined time period responsive to the detected cessation of the voice activity.

Ex. 1001, 13:1–25.

Petitioner relies on its arguments for claim 1 for limitations 13[preamble], 13[a], 13[c], 13[d], 13[f], 13[g], 13[h], 13[i], 13[j], and 13[l], and Patent Owner does not specifically address those limitations. Pet. 51–64; PO Resp. 67–69. Limitations 13[b], 13[e], and 13[j] will be discussed below.

For limitation 13[b], Petitioner argues that Rosenberg describes audio system 95 including "headphones (or other similar personalized audio presentation units that display audio content to the ears of a user) 95B," which one of ordinary skill would have understood to encompass "ear pieces that are placed/inserted in the ear canal." Pet. 51–52 (citing Ex. 1005 ¶ 40; Ex. 1002 ¶ 211). Petitioner argues that Kvaløy discloses a protective ear

terminal element inserted into an ear canal having a "sealing section 2" as part of the earplug which conforms to the shape of the user's ear canal. *Id.* at 52–53. Petitioner argues that one of ordinary skill would have been motivated to modify Rosenberg's ear pieces to include Kvaløy's sealing section configured to form to a user's ear canal to protect hearing and provide improved communication abilities in different noise environments. *Id.* at 53–54.

For limitation 13[e], Petitioner argues that one of ordinary skill would have recognized that Rosenberg's system would necessarily include a signal processing system in order to perform noise reduction, filtering, and/or other commonly known signal processing steps on the ambient signal, including voice detection. Pet. 55–56.

For limitation 13[j], Petitioner argues that Rosenberg does not specifically describe a "voice activity detector (VAD) timer system" to perform the claimed functions, but that Visser does so. Pet. 59–60. Specifically, Petitioner asserts that Visser describes an apparatus for processing audio signals that includes a voice activity detector configured to determine whether voice activity is present in segments of an audio signal, and also provides implementation details for using a voice activity detector to detect voice activity and process the audio in an earphone device. *Id.* at 59. Petitioner further argues that one of ordinary skill would have found it obvious to modify the Rosenberg-Kvaløy system's earphone to include a voice activity detector as described in Visser. *Id.* at 60.

Patent Owner argues that Rosenberg and Ichimura fail to disclose a delay responsive to a detected end of speech for the reasons set forth with respect to claim 1. PO Resp. 67–68. Petitioner also argues that Visser's

temporal smoothing renders Rosenberg's voice detection system inoperable for the reasons discussed with respect to claim 7, and Kvaløy's blocking feature is incompatible with Rosenberg for the reasons discussed with respect to claim 10. *Id.* at 68–69.

We agree with Petitioner's arguments. We find that the proposed combination discloses a delay responsive to a detected end of speech for the reasons discussed above with respect to claim 1. *See* § II.D.3(f), *supra.* We also find that it would have been obvious to combine Visser with Rosenberg, and that doing so would not render Rosenberg's voice detection system inoperable, for the reasons discussed above with respect to claim 7. *See* § II.E.2, *supra*. With respect to Kvaløy, Petitioner relies on Kvaløy's disclosure of a sealing section in the combination for claim 13, and we agree with Petitioner's arguments that it would have been obvious to combine Kvaløy's sealing section with the other references in the combination. Because Petitioner does not rely on Kvaløy's blocking function in the combination for claim 13, Patent Owner's arguments for claim 10 that the blocking function would interfere with Rosenberg's operation do not apply to claim 13.

Consequently, based on the full trial record, we agree with Petitioner's arguments, and find that Petitioner has proven by a preponderance of the evidence that claim 13 is unpatentable based on Rosenberg, Kvaløy, Visser, and Ichimura.

2. Dependent Claims 14–18

Claim 14 depends from claim 13 and further recites that "the at least one earphone device includes at least two earphone devices." Ex. 1001, 13:26–28. Petitioner argues that Rosenberg's disclosure of "ear pieces" is

sufficiently broad to encompass at least two ear pieces, and one of ordinary skill would have understood that earphones typically come in pairs, and users generally preferred audio systems with two headphones or ear pieces. Pet. 60–61.

Claim 15 depends from claim 13, and further recites that "the remote device includes at least one of a mobile phone, a radio device, a computing device, a portable media player, an earphone device of a different user or a further earphone device of the user." Ex. 1001, 13:29–33. Petitioner argues that Rosenberg discloses a "portable media player" and headphones that may be interfaced with the media player that is remote from the headphones via a wired or wireless connection. Pet. 61–62.

Claim 16 depends from claim 13 and further recites "a communication system configured to receive the AC signal from the remote device via a wired or wireless connection." Ex. 1001, 13: 34–37. Petitioner argues that Rosenberg discloses headphones that receive and audio representation of media content (an AC signal) output by a media player (a remote device) via a wired or wireless connection. Pet. 62.

Claim 17 depends from claim 13 and further recites that "the signal processing system is further configured to decrease the ASM gain of the ASM signal and increase the AC gain of the AC signal prior to mixing the ASM signal and the ASM signal when the voice activity is not detected." Ex. 1001, 13:38–14:2. Petitioner argues that Rosenberg in combination with Kvaløy, and Visser disclose or suggest this feature for the reasons discussed with respect to claims 2 and 13. Pet. 62.

Claim 18 depends from claim 13 and further recites "a voice activity detector (VAD) system configured to detect the voice activity from a

microphone signal, the microphone signal including at least one of the ASM signal or an ear canal microphone (ECM) signal captured within the ear canal from and ECM of the earphone device." Ex. 1001, 14:3–10. Petitioner argues that Rosenberg in combination with Kvaløy and Visser discloses or suggests the features of claim 18 for the reasons discussed with respect to claims 8 and 13. Pet. 63.

Patent Owner does not present separate argument for claims 14–18. PO Resp. 67–69.

Based on the full trial record, we agree with Petitioner's arguments, and find that the limitations of claims 14–18 would have been obvious over Rosenberg in view of Visser.

3. Dependent Claims 19–20

Claims 19 and 20 identical except that claim 19 depends from claim 18 and claim 20 depends from claim 19. Both claims recite that the VAD system is configured to "determine a time-smoothed level of the AC signal to form an AC level, compare the AC level with an AC level threshold, and detect the voice activity when the microphone level is greater than the microphone level threshold and the AC level is less than the AC threshold." Ex. 1001, 14:11–28.

Petitioner argues that Rosenberg in combination with Kvaløy and Visser disclose or suggest the features of these claims for the reasons discussed with respect to claims 10 and 18. Pet. 63. Patent Owner argues that this combination fails to teach claims 19 and 20 for the reasons discussed with respect to claim 10.

We agree with Patent Owner that the one of ordinary skill would not have combined Kvaløy with the other references as Petitioner proposes for

the reasons set forth above with respect to claim 10. PO Resp. 69; *see* § II.F.2, *supra*. Consequently, Petitioner has failed to prove that claims 19 and 20 are unpatentable based on the proposed combination.[9]

III. CONCLUSION

For the reasons discussed above, Petitioner has proven, by a preponderance of the evidence, that claims 1–9 and 11–18 are unpatentable, and failed to prove that claims 10, 19, and 20 are unpatentable, as summarized in the following table:[10]

Claim(s)	35 U.S.C. §	Reference(s)/Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1–6	103(a)	Rosenberg[11]		
1–6	103(a)	Rosenberg, Ichimura	1–6	
7–9	103(a)	Rosenberg, Visser	7–9	
7–9	103(a)	Rosenberg Lee[12]		
10	103(a)	Rosenberg, Visser, Kvaløy		10
13–20	103(a)	Rosenberg, Kvaløy, Visser[13]		

[9] For the same reason, Petitioner has failed to prove that claims 19 and 20 are unpatentable based on the combination of Rosenberg, Kvaløy, and Visser in Ground 4A.

[11] As explained above, because we find that claims 1–6 are unpatentable based on Rosenberg and Ichimura, we decline to address those claims in this ground.

[12] As explained above, because we find that claims 7–9 are unpatentable based on Rosenberg and Visser, we decline to address those claims in this ground.

[13] As explained above, because we find that claims 13–18 are unpatentable based on Rosenberg, Kvaløy, Visser, and Ichimura, we decline to address those claims in this ground.

13–20	103(a)	Rosenberg, Kvaløy, Visser, Ichimura	13–18	19, 20
Overall Outcome			1–9, 11–18	10, 19, 20

IV. ORDER

Accordingly, it is

ORDERED that claims 1–9 and 11–18 of the '542 patent have been proven by a preponderance of the evidence to be unpatentable; and

FURTHER ORDERED that claims 10, 19, and 20 of the '542 patent have not been proven by a preponderance of the evidence to be unpatentable; and

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

[13] As explained above, because we find that claims 13–18 are unpatentable based on Rosenberg, Kvaløy, Visser, and Ichimura, we decline to address those claims in this ground.

PETITIONER:

Ali R. Sharifahmadian
Jeffrey A. Miller
Mark Patrick
ARNOLD & PORTER KAYE SCHOLER LLP
ali.sharifahmadian@arnoldporter.com
jmillerptab@apks.com
mark.patrick @arnoldporter.com

PATENT OWNER:

Jacob A. Snodgrass
Bradford A. Cangro
Jeremy D. Peterson
PV LAW LLP
jacob.snodgrass@pvuslaw.com
bradford.cangro@pvuslaw.com
jeremy.peterson@pvuslaw.com
techiya-cases@pvuslaw.com