

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SONOS, INC.,
Petitioner,

v.

GOOGLE LLC,
Patent Owner.

IPR2022-01594
Patent 9,812,128 B2

Before KEVIN F. TURNER, TERRENCE W. McMILLIN, and
SCOTT RAEVSKY, *Administrative Patent Judges*.

McMILLIN, *Administrative Patent Judge*.

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

I. INTRODUCTION

A. *Background and Summary*

On September 29, 2022, Sonos, Inc. (“Petitioner”)¹ filed a Petition for *inter partes* review (“IPR”) of claims 1–3, 5–8, 10–13, and 15 (the “challenged claims”) of U.S. Patent No. 9,812,128 B2 (Ex. 1001, “the ’128 patent”). Paper 1 (“Pet.”). Google LLC (“Patent Owner”)² filed a Preliminary Response. Paper 6 (“Preliminary Response” or “Prelim. Resp.”). With our authorization, Petitioner thereafter filed a Reply to Patent Owner’s Preliminary Response (Paper 7) and Patent Owner filed a Sur-reply in Support of its Preliminary Response (Paper 8) to address the issue of discretionary denial under 35 U.S.C. § 325(d). Taking into account the arguments and evidence presented in these papers, we determined that the information presented in the Petition established that there was a reasonable likelihood that Petitioner would prevail with respect to challenging at least one of claims 1–3, 5–8, 10–13, and 15 of the ’128 patent as unpatentable. Pursuant to 35 U.S.C. § 314, we instituted this IPR on April 14, 2023, as to all challenged claims and all grounds raised in the Petition. Paper 9 (“Institution Decision” or “Dec. on Inst.”).

During trial, Patent Owner filed a Response (Paper 11, “PO Resp.”), Petitioner filed a Reply to the Response (Paper 14, “Pet. Reply”), and Patent Owner filed a Sur-reply to the Reply (Paper 18, “PO Sur-reply”). An oral argument was held on January 11, 2023, and a transcript of the hearing is included in the record. Paper 29 (“Tr.”).

¹ Petitioner identifies itself, Sonos, Inc., as the real party-in-interest to this proceeding. Pet. 1.

² Patent Owner identifies itself, Google LLC, as the real party-in-interest to this proceeding. Paper 4, 2.

We have jurisdiction under 35 U.S.C. § 6. This decision is a Final Written Decision under 35 U.S.C. § 318(a) as to the patentability of claims 1–3, 5–8, 10–13, and 15 of the ’128 patent. For the reasons we identify below, we hold that Petitioner has demonstrated by a preponderance of the evidence that all of the challenged claims are unpatentable.

B. Related Proceedings

Petitioner states that “[Patent Owner] asserted, *inter alia*, the ’128 Patent and [a related] U.S. Patent 11,024,311^[3] (the “’311 Patent”)[] against [Petitioner] in the following ITC and district court proceedings: *Certain Audio Players and Components Thereof*, Inv. No. 337-TA-1330;^[4] *Google LLC v. Sonos, Inc.*, No. 3:22-CV-04553 (N.D. Cal.) (“NDCA Action”).” Pet. 1–2, n.1 (formatting and footnote modified). Petitioner further states “[Patent Owner] has also asserted [related] U.S. Patents 10,134,398 and 10,593,330 against [Petitioner] in *Certain Audio Players and Components Thereof*, Inv. No. 337-TA-1329 and *Google LLC v. Sonos, Inc.*, No. 3:22-CV-04552 (N.D. Cal.).” Pet. 2. Patent Owner also identifies the above ITC proceeding and the litigation in the Northern District of California as “judicial and administrative matters that would be affected by[] a decision in this proceeding.” Paper 4, 2.

³ The ’311 patent is challenged in IPR2022-01592 in which the petition was filed on the same day as the Petition in this proceeding and in which the same art is asserted against the challenged claims.

⁴ The ITC Investigation was terminated as to the ’128 patent upon the motion of Patent Owner prior to the Initial Determination (*see* Ex. 1039, 2) and the Commission Opinion (*see* Ex. 1040, 4).

C. The '128 Patent

The '128 patent is titled “Device Leadership Negotiation Among Voice Interface Devices.” Ex. 1001, code (54). The patent “relates generally to voice interfaces and related devices, including but not limited to methods and systems for coordination amongst multiple voice interface devices.” *Id.* at 1:27–30. The described background regards “[e]lectronic devices integrated with microphones . . . [that] collect voice inputs from users and implement voice-activated functions according to the voice inputs . . . (e.g., Siri, Google Assistant).” *Id.* at 1:34–37. According to this description, “[a] location (e.g., a room or space within a home) may include multiple devices that include voice assistant systems and one or more users of such devices.” *Id.* at 1:43–45. And, “for a voice assistant device to be able to handle multiple users gracefully,” it is “desirable for there to be a leader amongst the voice assistant devices that . . . respond[s] to the user's voice inputs” and “for a voice assistant device to include noise mitigation measures . . . at a noisy location.” *Id.* at 1:47–55. In an “Example Device Leadership Negotiation” (*id.* at 20:43 (heading)):

[E]ach of the electronic devices determines a level of quality of the respective sample of the voice input captured by the electronic device. The quality level may be expressed as a score or some other value. . . . [C]riteria for evaluating the quality of a voice sample include, for example, loudness or amplitude of the sample, presence or absence of noise (e.g., cross-talk, hissing) and corresponding level, presence or absence of distortion (e.g., echo) and corresponding level, frequency profile of the sample, and so forth.

* * *

When an electronic device 1002 receives the scores from the other electronic devices 1002, the electronic device compares its quality score with the scores from the other

devices. If its score is not the best (e.g., highest) amongst the scores, the electronic device forgoes outputting responses . . . [and] forgo[s] receiving, detecting, and processing subsequent voice inputs from the user.

Id. at 21:1–4, 21:15–22, 21:58–22:10. Further, “[t]he leader determination is performed in response to a user’s voice input to awaken the device (e.g., a voice input that includes the hotword).” *Id.* at 22:11–13.

D. Challenged Claims

Petitioner challenges claims 1–3, 5–8, 10–13, and 15 of the ’128 patent. Pet. 1. The ’128 patent has three challenged independent claims: a method claim (claim 1), a device claim (claim 6), and a non-transitory computer-readable storage medium claim (claim 11). Ex. 1001, 35:54–36:14, 36:31–60, 37:11–38:8. Challenged dependent claims 2, 3, and 5 directly depend from claim 1; challenged dependent claims 7, 8, and 10 directly depend from claim 6; and challenged dependent claims 12, 13, and 15 directly depend from claim 11. *Id.* at 36:15–21, 36:25–30, 36:61–67, 37:4–10, 38:9–16, 38:20–28. Claim 1 recites:

1. A method, comprising:
at a first electronic device of a plurality of electronic devices,
each electronic device of the plurality of electronic devices
comprising one or more microphones, a speaker, one or more
processors, and memory storing one or more programs for
execution by the one or more processors:

detecting a voice input;

determining a quality score for the detected voice input;

communicating the quality score to the other devices of the
plurality of electronic devices;

receiving quality scores generated by the other devices of the
plurality of electronic devices for detection of the voice input
by the other devices;

in accordance with a determination that the quality score generated by the first electronic device is the highest amongst the generated quality score and received quality scores for the voice input, outputting an audible and/or a visual response to the detected voice input, wherein the other devices of the plurality of electronic devices forgo outputting an audible response to the detected voice input; and

in accordance with a determination that the quality score generated by the first electronic device is not the highest amongst the quality scores for the voice input generated by the plurality of electronic devices, forgoing outputting a response to the detected voice input.

Id. at 35:54–36:14 (emphasis added to disputed limitation). The disputed limitation (“detecting a voice input”) is recited identically in each of the challenged independent claims (*id.* at 35:61, 36:40, 37:18) and, by virtue of their dependency, is a limitation in every challenged dependent claim.

E. The Asserted Grounds

Petitioner challenges claims 1–3, 5–8, 10–13, and 15 of the ’128 patent based on the grounds set forth in the table below.

Ground	Claims Challenged	35 U.S.C.	Reference(s)
1	1–3, 5–8, 10–13, 15	§ 102	Rosenberger ⁵
2	5, 10, 15	§ 103	Rosenberger, Kim ⁶

Petitioner states that Rosenberger and Kim qualify as prior art respectively under 35 U.S.C. §102(a)(1) (AIA) & §102(a)(2) (AIA). Pet. 4. Patent Owner does not challenge the prior art status of Rosenberger and Kim. *See generally* PO Resp.

⁵ US 8,340,975 B1, published Dec. 25, 2012 (Ex. 1003).

⁶ US 2013/0238326 A1, published Sep. 12, 2013 (Ex. 1014).

Petitioner relies on the Declaration from Dr. Michael T. Johnson (Ex. 1002, “Johnson Decl.”) and Patent Owner relies on the Declaration of Stuart Lipoff (Ex. 2012, “Lipoff Decl.”).

II. ANALYSIS

A. Claim Construction

We construe claims using the same claim construction standard that would be used in a civil action under 35 U.S.C. § 282(b), as articulated by the Federal Circuit in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (*en banc*), and subsequent cases. *See* 37 C.F.R. § 42.100(b). Under the standard set forth in *Phillips*, claim terms are given their ordinary and customary meaning, as would have been understood by a person of ordinary skill in the art at the time of the invention, in light of the language of the claims, the specification, and the prosecution history of record. *See Thorner v. Sony Comput. Ent. Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012). “Absent lexicography or disavowal, we do not depart from the plain meaning of the claims.” *Luminara Worldwide, LLC v. Liown Elecs. Co. Ltd.*, 814 F.3d 1343, 1353 (Fed. Cir. 2016). In order for a patentee to act as its own lexicographer, it must “clearly set forth a definition of the disputed claim term” and “clearly express an intent to redefine the term.” *Thorner*, 669 F.3d at 1365 (internal quotation marks and citations omitted). Likewise, disavowal occurs only when there is a clear and unmistakable intent to limit the invention or exclude a particular feature from the claimed invention. *Luminara*, 814 F.3d at 1353. And, there is a “heavy presumption,” that a claim term carries its ordinary and customary meaning. *CCS Fitness, Inc. v. Brunswick Corp.*, 288 F.3d 1359, 1366 (Fed. Cir. 2002) (citation omitted).

In the Petition, Petitioner asserts that no claim construction is necessary. Pet. 23 (“For purposes of this IPR, the Challenged Claims do not require construction.”). In the Preliminary Response, Patent Owner does not address claim construction. *See generally* Prelim. Resp. Accordingly, in our Institution Decision, we determined that, “[a]t least with regard to making th[e] decision on institution . . . no claim construction [wa]s necessary.” Dec. on Inst. 8 (citing *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017) (explaining that construction is needed only for terms that are in dispute, and only as necessary to resolve the controversy)). Nonetheless, as explained below, claim construction and the meaning of the disputed limitation (“detecting a voice input”) has become an issue in this proceeding.

In its Response, Patent Owner raises the issue of claim construction. *See* PO Resp. 12–15. Patent Owner argues that “[t]he plain language of the claims and intrinsic evidence support that ‘detecting a voice input’/‘detect a voice input’^[7] as recited in independent claims 1, 6, and 11 refer to a process for identifying that an audio input includes voice.” *Id.* at 12. Patent Owner additionally argues in the Response that “detecting a voice input” should be construed as not including speech recognition. *Id.* The Response states that “[d]etecting’ a voice input and speech ‘recognition’ are distinct processes” and “[t]he claims, specification, and prosecution history support the understanding that ‘detecting a voice input’ refers to a process separate from performing speech recognition (e.g., speech to text) on the audio input.” *Id.* (citing Ex. 2012 at ¶¶ 69, 70). Patent Owner points to voice activity

⁷ “Detect a voice input” is not recited in independent claims 1, 6, or 11 of the ’128 patent. Ex. 1001, 35:54–36:14, 36:31–60, 37:11–38:8.

detection (VAD) and hotword/wakeword detection as processes that a skilled artisan would have understood as “detecting a voice input.” *Id.* at 14 (citing Ex. 2012 ¶ 71).

Patent Owner does not argue either lexicography or disavowal. *See* PO Resp. 12–15; PO Sur-reply 1–6. Petitioner also does not argue either lexicography or disavowal. Tr. 8:9–17. We do not find either (1) lexicography or (2) disavowal, because (1) there is no clearly set forth definition of “detecting a voice input” in the ’128 patent or (2) evidence of a clear and unmistakable intent to disavow the ordinary and customary meaning of “detecting a voice input.” Accordingly, we apply the ordinary and customary meaning of “detecting the voice input” in light of the language of the claims, specification, and file history (the intrinsic evidence). *Thorner*, 669 F.3d at 1365. In the Response and Reply, the parties assert different views of the ordinary and customary meaning of “detecting the voice input” as recited in the independent claims of the ’128 patent.

In the Reply, Petitioner argues that Patent Owner’s interpretation of “detecting a voice input” is “flawed.” Pet. Reply 7–20. Petitioner contends that Patent Owner’s “proposed construction contradicts the ’128 specification, confirming that it should be rejected.” *Id.* at 8. Petitioner asserts that, “in the ’128 Patent, the plain and ordinary meaning of ‘detecting a voice input’ is perceiving or sensing voice via a microphone.” *Id.* (citing Ex. 1038, ¶¶ 42–55). Petitioner argues “there is nothing in the claim language requiring ‘detecting a voice input’ to include anything beyond perceiving/sensing voice via a microphone” and “[m]oreover, the ’128 specification consistently describes the function of ‘detecting’ a voice input

interchangeably with the basic functions of ‘capturing’ or ‘receiving’ an utterance via a microphone.” *Id.* at 8–9 (citing Ex. 1001, 6:57–60, 24:3–6, 29:6–8; Ex.1038, ¶¶ 18, 44–47).

At the oral argument, both parties modified their positions on the meaning of “detecting a voice input” and agreed that “detecting a voice input” requires processing to identify that an audio input includes voice but does not exclude speech recognition. During the oral argument, Petitioner’s counsel said:

[Petitioner] would agree with -- okay, if you want to require some sort of identification of voice specifically, that’s fine. But you can’t go all the way where [Patent Owner] is asking to carve out speech recognition from the type of process that can be used to identify voice. That’s not permissible in [Petitioner’s] view.

* * *

[E]ven if we assume that you have to have some processing on the audio input to identify voice, there’s nothing in the intrinsic evidence that justifies the negative limitation that [Patent Owner] is seeking to add in.

Tr. 13:9–13, 18:22–25. Patent Owner’s counsel said:

And as Petitioner already said, they’re okay with the Board just taking that first part of the construction, that detecting a voice input requires just what it says: a process for identifying a voice input. [Patent Owner] would agree with that if that’s what’s needed to resolve the dispute, that that process at least requires that much.

If you don’t want it to go as far as to necessarily preclude speech recognition because the ITC^[8] determined that’s more a

⁸ Before the ITC, the parties proposed the same interpretations of “detecting a voice input” as used in the claims of the related ’311 patent which is based on a continuation of the ’128 patent (Ex. 2016, code (63)) as the parties proposed in this proceeding. *See* Ex. 1040, 12; *see also id.* at 17 n.18. The

question of whether the prior art lines up with speech recognition, then we would be okay with that.

I'm saying it's sufficient that if you would like to construe the term detecting a voice input to require just that, a process for identifying a voice input, that that is sufficient.

Our position has always been, and we'll show later on as we discuss Rosenberger, that this speech recognition doesn't read on detecting a voice input. And we can go into why. If you align with the ITC to the extent that that's really a question for what the prior art is disclosing on whether it matches the claim and it's not an issue for construction, we can just determine that the claim requires detecting a voice input, then that is okay. So I don't think it's a pivot in that we're changing the construction. Our position is the same. Rosenberger's speech recognition, his trigger phrase recognition, doesn't align with detecting a hotword, doesn't align with voice input detection. But if you want to construe the term to just be that and not reach the issue of speech recognition because that's a question that should be kept for the prior art, then that's okay

ITC determined “that ‘detecting a voice input’ as used in the claims and specification requires the device to identify the input as voice” and does not exclude speech recognition. *Id.* at 14. And, the ITC determined that “[w]hether an accused device—or, in this case, a prior art reference disclosing speech recognition—meets a limitation of the asserted claims is not a claim construction inquiry.” *Id.* at 15. We believe the analysis of the ITC as to the interpretation of “detecting a voice input” is well-reasoned and well-supported. We are in complete agreement with the ITC. And, Patent Owner acknowledges that the independent claims of the ’311 patent that is the subject of the ITC’s analysis are substantially similar to the independent claims of the ’128 patent. PO Resp. 13. We believe that the ITC’s analysis and conclusions regarding interpretation of “detecting a voice input” are equally applicable to the claims of the ’128 patent.

Id. at 29:23–30:2, 30:12–15, 30:23–25, 31:3–15.⁹ Thus, the parties have agreed that the ordinary and customary meaning of “detecting a voice input” requires a process for identifying that an input includes voice and does not exclude using speech recognition as the process to identify that the audio input includes voice. For the reasons discussed below, we accept requiring a process for identifying that an input includes voice and not excluding speech recognition as the ordinary and customary meaning of “detecting a voice input.”

In the “Claim Construction” section of the Response (PO Resp. 12–15), Patent Owner cites the following passages from the ’128 patent:

[A] “hotword” is a user defined or predefined word or phrase used to “wake-up” or trigger a voice-activated device to attend/respond to a spoken command that is issued subsequent to the hotword) or other spoken voice inputs.

* * *

Hotword recognition module 562 for determining whether voice inputs include a hotword for waking up the electronic device 190 and recognizing such in the voice inputs;
Request recognition module 564 for determining a user request included in a voice input;

* * *

The user makes a voice input (e.g., an utterance) (not shown), and the three electronic devices 1002-1 thru 1002-3 detect and capture a sample of the voice input at varying degrees or levels of quality.

* * *

When an electronic device 1002 receives the scores from the other electronic devices 1002, the electronic device

⁹ At oral argument, Patent Owner’s counsel confirmed its position that hotword detection is a process for “detecting a voice input” as recited. Tr. 31:19–26.

compares its quality score with the scores from the other devices. If its score is not the best (e.g., highest) amongst the scores, the electronic device forgos [sic] outputting responses (e.g., responses generated by the server system 140) to the voice input (e.g., by powering down or “turning off” the speaker 520). If its score is the best amongst the scores, the electronic device outputs responses to the voice input and subsequent voice inputs; this device is determined to be the “leader.”

Ex. 1001, 5:55–58, 17:8–13, 20:60–62, 21:58–22:1. In its Response, Patent Owner argues that these passages establish that “[t]he specification . . . distinguishes between speech recognition and hotword detection (or other methods of detecting a voice input).” PO Resp. 14. However, we find nothing in these passages from the Specification of the ’128 patent that supports the distinction between speech recognition and hotword detection that Patent Owner argues. To the contrary, the second passage states that the “Hotword recognition module . . . determin[es] whether voice inputs include a hotword for waking up the electronic device . . . and recognizing such in the voice inputs” (Ex. 1001, 17:8–11) which provides that hotword detection is a process that recognizes voice inputs or, stated differently, performs speech recognition. And, we find nothing in these passages that supports attributing a meaning to “detecting a voice input” that excludes speech recognition. Thus, Patent Owner’s proposed negative limitation finds no support in the Specification of the ’128 patent.

Patent Owner also argues that claim differentiation supports its position that “detecting a voice input” refers to a process separate from speech recognition. PO Resp. 12–14. Specifically, Patent Owner contends “that ‘detect[ion],’ as recited in claim 6, is a separate process from

‘recogni[tion],’ as recited in claim 10.” *Id.* at 12. As discussed above, independent claim 6 recites “detecting a voice input.” Ex. 1001, 36:40. Claim 10 depends from claim 6 and recites:

10. The device of claim 6, further comprising instructions for:
recognizing a command in the voice input; and
in accordance with a determination that a type of the command is related to the first electronic device outputting an audible and/or a visual response to the detected voice input.

Id. at 37:4–10 (emphasis added). The relevant limitations in these claims are specifically referring to “programs comprising instructions.” *Id.* at 36:38–39. But, because these claims recite that different instructions are used for “detecting a voice input” (i.e., determining whether the input includes voice or speech) and for “recognizing a command in the voice input,” this claim language does not support Patent Owner’s contention that “detecting a voice input” excludes speech recognition. Patent Owner does not adequately explain, and we do not discern, why both recited processes or sets of instructions cannot use speech recognition processes and instructions.¹⁰ We do not find Patent Owner’s claim differentiation argument to have merit.

As noted above, Petitioner contends that “in the ’128 Patent, the plain and ordinary meaning of ‘detecting a voice input’ is perceiving or sensing voice via a microphone.” Pet. Reply 8 (citing Ex. 1038 ¶¶ 42–55).

¹⁰ Patent Owner’s similar argument was rejected by the ITC. *See* Ex. 1040, 19 (“[V]oice activity detection [(VAD)] may be sufficient to satisfy the ‘detecting a voice input’ limitation as used in the ’311 patent, but does not preclude the speech recognition function of an allegedly infringing device or prior art reference from meeting that limitation as well.”).

Petitioner cites the following passages from the '128 patent in support of its interpretation of “detecting a voice input” (*id.* at 8–10):

When a user 102 makes a voice input, each of the devices 190 either receives the input or does not receive the input (e.g., if the device was too far away from the user).

* * *

In some implementations, if just one of the multiple electronic devices 1002 captured a sample of the voice input (e.g., the other electronic devices 1002 were too far away from the user to detect the voice input), the device that captured the voice input sample is determined to be the leader.

* * *

The electronic device 190 receives (1102) via the one more microphones 516 a first voice input from a user. The electronic device 190 captures and records a sample of the first voice input via the microphone 516.

* * *

The electronic device 190 detects (1202) a voice input. The electronic device 190 captures and records a sample of the voice input from a user via the microphone 516.

Ex. 1001, 6:57–60, 21:47–52, 24:3–6, 29:6–8. We do not dispute that, without the microphone or some similarly functioning component, the device could not detect the input. However, Petitioner’s interpretation would be equally applicable if the limitation were “detect an input” and fails to give meaning to the modifier “voice” in the phrase “detecting a voice input” as recited. The abstract of the '128 patent explains that the device “detect[s] a voice input” and “determin[es] a quality score for the detected voice input.” Ex. 1001, code (57). Determining a quality score for the detected voice input implies some level of identification of the input as a voice, and not mere perceiving or sensing by the microphone. And, the

Specification states that “in accordance with some implementations, a voice-activated device determines whether the surrounding noise is sufficiently interfering with detection of spoken hotwords for waking up the device . . . or other spoken voice inputs.” *Id.* at 5:51–58. Thus, the ’128 patent discloses that the device distinguishes between background noise and the detection of specific voice inputs like hotwords. We conclude that the ordinary and customary meaning of “detecting a voice input” in the context of the claims and Specification of the ’128 patent requires the device to identify the input as voice.¹¹

In sum, we conclude that the ordinary and customary meaning of “detecting a voice input” requires a process for identifying that an audio input includes voice and does not exclude speech recognition.

B. Legal Standards

In order for a prior art reference to serve as an anticipatory reference, it must disclose every limitation of the claimed invention, either explicitly or inherently. *In re Schreiber*, 128 F.3d 1473, 1477 (Fed. Cir. 1997).

Anticipation “requires that every element and limitation of the claim was previously described in a single prior art reference, either expressly or inherently, so as to place a person of ordinary skill in possession of the invention.” *Sanofi-Synthelabo v. Apotex, Inc.*, 550 F.3d 1075, 1082 (Fed. Cir. 2008) (citing *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1379 (Fed. Cir. 2003); *Continental Can Co. USA v. Monsanto Co.*, 948 F.2d 1264, 1267–69 (Fed. Cir. 1991)). “[U]nless a reference discloses within the

¹¹ Here again, we note that, in response to similar arguments, the ITC rejected Petitioner’s interpretation and concluded “that ‘detecting a voice input’ as used in the claims and specification [of the related ’311 patent] requires the device to identify the input as a voice.” Ex. 1040, 14.

four corners of the document not only all of the limitations claimed but also all of the limitations arranged or combined in the same way as recited in the claim, it cannot be said to prove prior invention of the thing claimed and, thus, cannot anticipate under 35 U.S.C. § 102.” *Net MoneyIN, Inc. v. VeriSign, Inc.*, 545 F.3d 1359, 1371 (Fed. Cir. 2008).

A patent claim is unpatentable as obvious if the differences between the claimed subject matter 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 ordinary skill in the art; and (4) when presented, objective evidence of non-obviousness. *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966).

“In an [*inter partes* review], the petitioner has the burden from the onset to show with particularity why the patent it challenges is unpatentable.” *Harmonic Inc. v. Avid Technology, Inc.*, 815 F.3d 1356, 1363 Fed. Cir. 2016) (citing 35 U.S.C. § 312(a)(3) (requiring *inter partes* review petitions to identify “with particularity . . . the evidence that supports the grounds for the challenge to each claim”)). Petitioners cannot satisfy their burden of proving obviousness by employing “mere conclusory statements.” *In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1380 (Fed. Cir. 2016).

C. Level of Ordinary Skill in the Art

With regard to the level of ordinary skill in the art, Petitioner contends a “POSITA [(person of ordinary skill in the art)] in the technology field of the ’128 Patent would have had the equivalent of a four-year degree from an accredited institution (typically denoted as a B.S. degree) in electrical or computer engineering, or a related field, and approximately 2–4 years of experience in the field of voice and/or audio processing, or an equivalent level of skill, knowledge, and experience.” Pet. 22 (citing Ex. 1002 ¶¶ 63–65.). Patent Owner does not present a position on the level of ordinary skill in the art at the time of the invention. *See generally* PO Resp.

Petitioner’s proposal is consistent with the technology described in the Specification of the ’128 patent and the cited prior art. We adopt Petitioner’s proposed level of skill in the art.

D. Challenges Based on Rosenberger

Petitioner contends that claims 1–3, 5–8, 10–13, and 15 are anticipated by Rosenberger and that claims 5, 10, and 15 are also rendered obvious by a combination of Rosenberger and Kim. Pet. 4. The Petition includes a detailed element-by-element mapping of the limitations in the challenged claims to teachings in the cited art and supports the mappings with references to Dr. Johnson’s Declaration. Pet. 23–69. Except with regard to the “detecting a voice input” limitation, Patent Owner does not dispute Petitioner’s showings as to the claim limitations. *See generally* PO Resp.

In our Institution Decision, we considered Petitioner’s showing that Rosenberger discloses all the limitations of the challenged claims of the ’128 patent. *See* Dec. on Inst. 14–30. We determined that Petitioner’s showing

that Rosenberger disclosed the limitations of the challenged claims was well-supported (*see id.*) and, with specific regard to claim 1, we noted that the challenge to claim 1 as anticipated by Rosenberger appeared to be strong (*see id.* at 25). We have reconsidered Petitioner’s showing in view of the preponderance of evidence standard applicable to this final written decision. *See* 35 U.S.C. § 316(e). We find based upon Petitioner’s showing and the evidence cited (*see* Pet. 23–69) and based on our analysis as set forth in the Institution Decision that all the limitations of the challenged claims (including the “detecting a voice input” limitation) are disclosed in Rosenberger. We conclude that Rosenberger anticipates claims 1–3, 5–8, 10–13, and 15 of the ’128 patent.

In light of Patent Owner’s dispute of Petitioner’s showing as to the “detecting a voice input” limitation, we discuss the arguments and evidence of the parties as to that limitation below. We begin our analysis with a summary of the teachings of Rosenberger.

1. Rosenberger (Ex. 1003)

Rosenberger is titled “Interactive Speech Recognition Device and System for Hands-Free Building Control.” Ex. 1003, code (54). Rosenberger relates to “relates to improvements in automated control systems” and, “[m]ore particularly, . . . to improvements in speech-recognizing control devices that respond to oral commands from a system user.” *Id.* at 1:8–13.

The described prior art regards speech-recognition for waking such systems; e.g., a “speech password mode . . . constantly listening for one or more passwords . . . referred to as ‘trigger phrases’” and a “circle recognition mode . . . constantly listening for a complete set of allowable

command phrases.” *Id.* at 2:4–11. According to the description, “a single control device may not be capable of reliably recognizing a speech command or trigger phrase from all locations throughout the given room” due to “ambient noise, room size, microphone quality, device location, room acoustics, furniture positions and a variety of other factors.” *Id.* at 2:26–30. For “modules in reasonable proximity to one another, [e.g.,] several per room, . . . a given hands-free speech command would frequently be recognized simultaneously by two or more neighboring speech recognition modules . . . with potentially undesirable or adverse effects” such as a “duplicate response” (e.g., twice lowering a thermostat). *Id.* at 2:45–60.

A summary of the disclosed solution is:

According to a preferred embodiment, the control device of the invention includes (a) a microphone for converting spoken words, emanating from a system user located within the building space, to corresponding electrical signals; (b) a programmable microcontroller, operatively coupled to the microphone and comprising embedded software and circuitry serving to (i) recognize from the electrical signals when a speech command from an expected vocabulary or grammar has likely been spoken, and (ii) to act thereon to first produce a weighted signal representing the quality of a just-recognized speech command, and then produce, only in the event the value of said weighted signal exceeds the respective values of all weighted signals produced by other devices within said building space that simultaneously recognized the same speech command, predetermined system control signals by which a desired effect, reflective of the speech commands, is achieved within the building space by system controllers or automated mechanisms that respond to such system control signals; (c) a transmitter operatively coupled to the microcontroller for selectively transmitting the weighted signal and the predetermined system control signals into the building space; (d) a receiver, operatively coupled to the microcontroller, for receiving weighted signals emanating from one or more other

control devices within the building space; and (e) audio output circuitry through which the microcontroller can transmit audible sounds to interact with and provide indications of the system status to the system user.

A preferred characteristic of the control device of the invention is that, within a predetermined time interval after it has produced the weighted signal in response to recognizing a given speech command, the microcontroller operates to compare the value of this signal with the respective value(s) of the weighted signals produced by all other control devices that have been triggered by the same speech command at the same time and have transmitted their respective weighted signals to such microcontroller. In the event the value of such weighted signal exceeds the respective values of all other weighted signals with which it has been compared, the microcontroller proceeds in handling the given speech command. Otherwise, it defers handling of the speech command to the other devices and reverts to a “listening” mode awaiting subsequent speech commands.

Id. at 4:6–47.

Rosenberger issued on December 25, 2012, based on an application filed on October 4, 2011. Ex. 1003, codes (22), (45). The earliest priority date claimed for the ’128 patent is based on a provisional application filed on October 9, 2014. Ex. 1001, code (60). Rosenberger is prior art to the ’128 patent.

2. Analysis Relating to the “Detecting a Voice Input” Limitation

Petitioner relies on Rosenberger’s disclosure of a device that receives and recognizes a spoken trigger phrase as disclosing “detecting a voice input” as recited. *See* Pet. 27–30 (citing Ex. 1003, 1:53–2:9, 4:6–14, 8:17–24, 9:57–61, 14:10–18, Fig. 4). The first cited passage is under the headings “Background of the Invention” and “The Prior Art” in Rosenberger and states:

Recently, another hands-free automated home control system has been proposed that responds to speech commands from a system user to reduce the need for physical contact with the systems user interface. Such a system is described in a paper entitled “Developing a Voice Control System for ZigBee Home Automation Networks.” published in the Proceedings of IC-NIDC2010, by researchers Jieming Zhu et al. at Beijing University. Note, ZigBee is a registered trademark of ZigBee Alliance and refers to a low-cost, low-power wireless RF (radio frequency) mesh networking standard that is widely used in wireless building control and monitoring applications. Here, a voice-controlled home automation system combines one or more speech recognition modules with conventional ZigBee-based wireless sensor and actuator networks to enable a system user to wirelessly control the operation of automated mechanisms via speech commands. Two different modes of hands-free operation are discussed in this paper, i.e. a “speech password mode” and a “circle recognition mode”. In the speech password mode, a speech recognizer is constantly listening for one or more passwords, commonly referred to as “trigger phrases”. When a trigger phrase is detected, the system “wakes up” and prompts the user to say one or more commands from an expected command phrase vocabulary.

Ex. 1003, 1:53–2:9. The other cited passages in Rosenberger provide:

According to a preferred embodiment, the control device of the invention includes (a) a microphone for converting spoken words, emanating from a system user located within the building space, to corresponding electrical signals; (b) a programmable microcontroller, operatively coupled to the microphone and comprising embedded software and circuitry serving to (i) recognize from the electrical signals when a speech command from an expected vocabulary or grammar has likely been spoken, and (ii) to act thereon.

* * *

Preferably, device D normally operates in a low power “listening” mode in which it listens for a hands-free trigger phrase (e.g., “Hello Voice Control” or “Home Automation”) to

be spoken by the system user. Upon receiving and recognizing a speech trigger phrase and determining that it is in a better position to handle subsequent user interaction than any other device that simultaneously recognized the same speech trigger phrase (see device coordination discussion below).

* * *

Still referring to FIG. 3 [a residential floor plan], if person P3 stands in the dining room near the kitchen island and speaks a trigger phrase such as “Hello Voice Control”, table top devices T2, T4, T5 and T6 and in-wall devices W5 and W9 may all successfully “hear” and recognize this speech trigger.

* * *

In FIG. 7 [block diagram of illustrative system architecture], a building automation system S is schematically illustrated as comprising a plurality of the speech recognition devices (D1-D4) of the invention in combination with a coordinating controller CC of the type described above. The coordinating controller receives (wirelessly or via power line communication) the weighted signal messages WS from the control devices and, based on the respective values of these signals, selects one device to handle a speech command (or Trigger Phrase) heard by at least two control devices.

Id. at 4:6–14, 8:17–24, 9:57–61, 14:10–18. Cited Figure 4 is a flowchart in which the first step is “Speech Trigger or Command Recognized.”¹² In these passages, Rosenberger discloses a method and device utilizing embedded software and circuitry that receives and recognizes a spoken trigger phrase (such as “Hello Voice Control”) and takes action based thereon. Thus, Rosenberger discloses a process for identifying that an audio input includes voice or “detecting a voice input” as recited.

¹² Figures 5A and 6 are also flowcharts with “Speech Trigger or Command Recognized” steps.

Petitioner also contends that during prosecution of a related application that a Patent Examiner asserted that Rosenberg teaches “detecting a voice input.” Pet. 29 (citing Ex. 1005, 61–62 (“Rosenberger teaches . . . A computer-implemented method comprising . . . determining, by the computing device, that the audio data likely includes an utterance of a particular, predefined hotword.”)). And, Petitioner contends that Patent Owner conceded that Rosenberg discloses this limitation during prosecution of a related application. *Id.* at 30 (citing Ex. 1005, 78–79 (“Rosenberger discloses that when multiple devices T6, T7, W8, W9 within an environment all recognize a same trigger phrase, ‘Hello Voice Controller’, each device will calculate a ‘weighted signal’ and communicate the weighted signal to a coordinating controller.”)).

In its Response, Patent Owner provides a “Background of the Technology” section. *See* PO Resp. 1–11. In this “Background” section, Patent Owner includes a description of the stages in which voice assistants process audio that is called a pipeline. *Id.* at 1. One of the pipeline stages that Patent Owner describes is a “Detection (Voice Activity Detection / Hotword Detection / Wakeword Detection)” stage. *Id.* at 4–8. With specific regard to hotword/wakeword detection, Patent Owner’s description is “the detection of particular words or phrases (e.g., ‘Hey Google’) in an audio signal that a user speaks to initiate an interaction with the voice assistant.” *Id.* at 7. Patent Owner’s description of hotword/wakeword detection fits the process disclosed in the above-quoted passages from Rosenberg. And, as noted above, at oral argument, Patent Owner’s counsel confirmed its position that hotword detection is a process for ‘detecting a voice input’ as recited. Tr. 31:19–26.

In its Response, Patent Owner argues that Rosenberg does not disclose this limitation based on a distinction between detection and recognition, such as speech recognition. PO Resp. 15–24. Patent Owner argues that “Rosenberger explains that the embedded speech recognizer returns a ‘recognition confidence’ for trigger phrases which requires performing speech recognition.” *Id.* at 20 (citing Ex. 1003, 10:41–52). But, as discussed above, Patent Owner, at oral argument, abandoned its contention that “detecting a voice input” excludes speech recognition. *See* Tr. 29:23–30:2, 30:12–15, 30:23–25, 31:3–15. And, the sentence immediately preceding the passage cited in Rosenberg states that “[a]ll of the methods involve calculation of a numeric weighted signal (WS) that quantifies the ‘quality’ of a just-recognized command.” Ex. 1003, 10:38–41. This sentence further supports a finding that Rosenberg discloses a method and devices that detect and recognize a spoken audio input and process it by quantifying it.¹³ We have considered Patent Owner’s arguments and the

¹³ Independent claim 19 of Rosenberg is directed to devices including:

- (a) a microphone for converting spoken words, emanating from a system user located within said building space, to corresponding electrical signals;
- (b) a programmable microcontroller, operatively coupled to the microphone and comprising embedded software and circuitry serving to:
 - (i) continuously monitor said electrical signals in order to recognize when a trigger phrase from an expected vocabulary has likely been spoken,
 - (ii) produce a weighted signal having a value representing the quality of a just-recognized trigger phrase.

Ex. 1001, 20:1–13. And, dependent claim 22 of Rosenberg recites that the “weighted signal incorporates one or more quantifiable measures

distinction it attempts to draw between detecting and recognizing and find it unavailing. Moreover, the disclosure of recognition of a vocal input strongly supports finding that the vocal input was detected.

E. Analysis Relating to Obviousness of Claims 5, 10, and 15 in View of Rosenberger and Kim

Petitioner also challenges claims 5, 10, and 15 of the '128 patent as obvious over the combination of Rosenberger and Kim. Pet. 4, 56–69. The additional limitations recited in dependent claims 5, 10, and 15 are identical. Ex. 1001, 36:25–30, 37:4–10, 38:20–28. The additional limitations in claims 5, 10, and 15 are:

recognizing a command in the voice input; and
in accordance with a determination that a type of the command
is related to the first electronic device, outputting an audible
and/or a visual response to the detected voice input.

Id. Petitioner contends that both Rosenberger and Kim disclose the additional limitations of claims 5, 10, and 15. Pet. 52–62. In our Institution Decision, we determined that Rosenberger discloses the additional limitations of claims 5, 10, and 15. Dec. on Inst. 29–30.¹⁴ Accordingly, in

representing the detected trigger phrase loudness and/or the estimated probability that the trigger phrase was correctly recognized by the embedded software of the microcontroller.” *Id.* at 21:1–5. These claims provide further support for our finding that Rosenberger discloses devices that detect and recognize voice input (words spoken by a user) and process this voice input by quantifying it and that Rosenberger discloses the “detecting a voice input” limitation.

¹⁴ We considered Petitioner’s assertion that Rosenberger anticipated claims 5, 10, and 15 in our Institution Decision. Dec. on Inst. 29–30. As indicated above, we have reconsidered Petitioner’s showing as to anticipation in view of the preponderance of evidence standard and find Rosenberger discloses all elements of the limitations of claims 5, 10, and 15 and anticipates these claims.

our analysis below, we consider whether Kim also teaches or suggests the additional limitations of claims 5, 10, and 15, and motivation to combine the relevant teachings of Rosenberger and Kim.

Patent Owner does not dispute that Kim teaches the additional limitations of claims 5, 10, and 15. *See* PO Resp. 29–30. With regard to “recognizing a command in the voice input,” Petitioner cites the following passages in Kim (Pet. 56–57):

To achieve this objective of the present invention, an aspect is directed to a method of recognizing a voice command by a device, the method comprising: receiving a voice input; processing the voice input by a voice recognition unit, and identifying at least a first voice command as including attribute information corresponding to the device from the voice input; recognizing the first voice command as being intended for the device based on at least the attribute information corresponding to the device identified from the first voice command, and controlling the device according to the recognized first voice command.

* * *

The voice recognition unit 103 [of Figure 1] receives a user's voice input that is picked up by the microphone 104 and performs a voice recognition process on the audio data corresponding to the user's voice input in order to interpret the meaning of the user's voice input. The voice recognition unit 103 may then perform processing on the interpreted voice input to determine whether the voice input included a voice command intended to control a feature of the voice recognition capable device 100. A more detailed description for the voice recognition processing accomplished by the voice recognition unit 103 will be provided throughout this disclosure.

Ex. 1014 ¶¶ 8, 48. Petitioner’s undisputed contention that Kim teaches this limitation is well-supported. We find Kim teaches this limitation of claims 5, 10, and 15.

With regard to “in accordance with a determination that a type of the command is related to the first electronic device, outputting an audible and/or a visual response to the detected voice input” as recited in claims 5, 10, and 15, Petitioner provides a showing that all the elements of this limitation are taught by Kim. Pet. 57–62 (citing Ex. 1002 ¶¶ 174–182; Ex. 1014 ¶¶ 8, 51–54, 58, 61, 67–69, 72–74, 78, Fig. 7). In the Petition, this illustrative example from Kim is used to show Kim teaches this limitation (*id.* at 58–61):

The following provides an example where a volume setting feature may be an attribute that is supported to be controlled by a voice command, for example, on a voice recognition capable device. When a user announces a voice command for controlling a volume setting in the presence of the television 210, mobile communication device 220, laptop computer 230 and refrigerator 240 in the environment illustrated by FIG. 2, each of these voice recognition capable devices may receive/hear the user's voice command. Then the voice recognition unit 103 for each respective voice recognition capable device will process the user's voice command and identify the volume feature as the attribute included in the voice command. After identifying the volume feature as the attribute that is intended to be controlled by the user's voice command, only the television 210, mobile communication device 220, laptop computer 230 may actually recognize the voice command as potentially being intended for it because only these voice recognition capable devices are capable of supporting a volume setting attribute. This is because the television 210, mobile communication device 220, laptop computer 230 inherently support a volume setting feature. Because the refrigerator 240 (in most cases) is not capable of supporting the volume setting attribute, the refrigerator 240 may hear the user's volume setting voice command but it will not recognize the volume setting voice command as intended for it after identifying the volume setting as the attribute from the user's voice command.

To narrow things even further, in some embodiments of the present invention, a voice recognition capable device may not recognize a user's voice command if the attribute identified from the user's voice command is not currently being utilized by the voice recognition capable device. This is true even if the voice recognition capable device inherently supports such an attribute. For instance, if the mobile communication device 220 and the laptop computer 230 are not specifically running an application that requires a volume setting when the user's volume setting voice command is announced, then if the television 210 is currently displaying a program, then the television 210 may be the only device from amongst the plurality of devices to recognize the volume setting voice command and perform a volume setting control in response to the user's volume change voice command. This additional layer of smart processing offered by the present invention provides a more accurate prediction of determining the true intention of a user's voice command.

Ex. 1014 ¶¶ 52–53. Petitioner’s undisputed contention that Kim teaches this limitation is well-supported. We find Kim teaches this limitation of claims 5, 10, and 15.

Petitioner contends that a skilled artisan “would have been motivated to modify Rosenberger to incorporate Kim’s teachings such that Rosenberger’s ‘control device’ would respond to a detected voice input in accordance with a determination that a type of a command recognized in the voice input is related to the ‘control device’ . . . for several reasons.” Pet. 63. The articulated reasons supporting the alleged motivation to combine include:

Rosenberger and Kim (like the ’128 Patent) are in the same field of endeavor and relate to voice-controllable electronic devices and provide solutions for intelligently processing a voice input detected by multiple electronic devices in a system.

See, e.g., Ex.1003, 2:41-3:26, 9:18-10:41; Ex.1014, ¶¶1-5, 50-51; Ex.1002, ¶185.

A POSITA would have been motivated to incorporate the teachings of Kim into Rosenberger’s “control device” in view of Rosenberger’s goal of “selecting **a single control device best able** to successfully interact with the system users.” Ex.1003, 10:34-38; Ex.1002, ¶186.

* * *

[A] POSITA would have been motivated to incorporate Kim’s voice-command attribute analysis into Rosenberger’s “control device” because it would have simply involved applying a known technique (i.e., voice-command attribute analysis) to a known device (i.e., Rosenberger’s voice-enabled “control device”) that was ready for improvement to yield predictable results (i.e., Rosenberger’s “control device” being able to respond to a detected voice input in accordance with a determination that a type of a command recognized in the voice input is related to the “control device”). Ex.1002, ¶190.

Id. at 63, 65–66. We find Petitioner’s articulated reasons for combining the relevant teachings of Rosenberger and Kim to be well-supported.

The Response contains this single sentence criticizing Petitioner’s showing of motivation to combine: “[A]ll of the Petition’s motivations to combine are clearly rooted in hindsight because, in Rosenberger, only devices that recognize a particular command (i.e., the devices that a voice command is intended for in the disclosure of Kim) need to participate in the arbitration process.” PO Resp. 30 (citing Ex. 2012 ¶ 87). This argument is not developed or explained and it contradicts the disclosure of Rosenberger. In Rosenberger, all the devices that detect and recognize a voice input (the trigger phrase) participate in the coordination or arbitration process. *See, e.g.*, Ex. 1003, 9:57–10:33. This argument lacks merit.

We find that the combination of Rosenberger and Kim teaches all the limitation of claims 5, 10, and 15 of the '128 patent and that a motivation to combine the relevant teachings of these references existed. We conclude that a clear preponderance of the evidence establishes that claims 5, 10, and 15 of the '128 patent are unpatentable as obvious.

F. Summary of Analysis

The clear preponderance of the evidence supports Petitioner's contention that Rosenberger discloses the "detecting a voice input" limitations of the challenged claims of the '128 patent. This is the only limitation that Patent Owner argues. *See generally* PO Resp.; PO Sur-reply. Patent Owner relies on its arguments relating to this limitation for both grounds asserted in the Petition. *See* PO Resp. 29–30. The single sentence in the "Conclusion" of the Response states, "[f]or the aforementioned reasons, Petitioner fails to meet its burden of establishing that the claims are unpatentable as anticipated by Rosenberger." *Id.* at 30. We disagree and conclude, based on the foregoing analysis, that a preponderance of the evidence relied on in the Petition establishes that all of the challenged claims are anticipated by Rosenberger and claims 5, 10, and 15 are rendered obvious by the combination of Rosenberger and Kim.

III. CONCLUSION¹⁵

Petitioner has demonstrated by a preponderance of the evidence that claims 1–3, 5–8, 10–13, and 15 of the '128 patent are unpatentable. A summary of our conclusions is set forth in the table below.

Claims	35 U.S.C. §	Reference(s)/Basis	Claims Shown Unpatentable	Claims Not shown Unpatentable
1–3, 5–8, 10–13, 15	102	Rosenberger	1–3, 5–8, 10– 13, 15	
5, 10, 15	103	Rosenberger, Kim	5, 10, 15	
Overall Outcome			1–3, 5–8, 10– 13, 15	

IV. ORDER

In consideration of the foregoing, it is

ORDERED that claims 1–3, 5–8, 10–13, and 15 of U.S. Patent No. 9,812,128 B2 are held to be unpatentable; and

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

¹⁵ Should Patent Owner wish to pursue amendment of the challenged claims in a reissue or reexamination proceeding subsequent to the issuance of this decision, we draw Patent Owner's attention to the April 2019 *Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding*. See 84 Fed. Reg. 16,654 (Apr. 22, 2019). If Patent Owner chooses to file a reissue application or a request for reexamination of the challenged patent, we remind Patent Owner of its continuing obligation to notify the Board of any such related matters in updated mandatory notices. See 37 C.F.R. § 42.8(a)(3), (b)(2).

IPR2022-01594
Patent 9,812,128 B2

FOR PETITIONER:

Michael Boyea
Jaeyoung Pak
Cole Richter
LEE SULLIVAN SHEA & SMITH LLP
boyea@ls3ip.com
pak@ls3ip.com
richter@ls3ip.com

FOR PATENT OWNER:

Erika Arner
Cory Bell
Alissa Green
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER LLP
erika.arner@finnegan.com
cory.bell@finnegan.com
alissa.green@finnegan.com