

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SONOS, INC.,
Petitioner,

v.

GOOGLE, LLC,
Patent Owner.

Case No. IPR2023-01288
U.S. Pat. No. 9,632,748

PATENT OWNER'S PRELIMINARY RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	THE EVIDENCE AND ITC CONFIRM THAT ROSENBERGER DOES NOT DISCLOSE DETERMINING WHETHER TO PEFORM SPEECH RECOGNITION	2
A.	Rosenberger’s Control Device Performs Speech Recognition Before Information Is Received from Other Devices	2
B.	The ITC Agrees that Rosenberger Does Not Disclose Performing Speech Recognition Based on a Determination, as Claimed.....	14
III.	CONCLUSION.....	15

TABLE OF AUTHORITIES

	Page(s)
Federal Cases	
<i>Microsoft Corp. v. Biscotti, Inc.</i> , 878 F.3d 1052 (Fed. Cir. 2017)	3
<i>Net MoneyIN, Inc. v. VeriSign, Inc.</i> , 545 F.3d 1359 (Fed. Cir. 2008)	4
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.</i> , 851 F.3d 1270 (Fed. Cir. 2017)	4

EXHIBIT LIST

Exhibits	Description
Ex. 2001	<i>Certain Audio Players and Components Thereof (II)</i> , Inv. No. 337-TA-1330 (“1330 ITC Investigation”), Initial Determination on Violation of Section 337 and Recommended Determination on Remedy and Bond (Sept. 15, 2023)

I. INTRODUCTION

Petitioner asserts that claims 1-4, 7, 9-12, 14, and 15 (of which claims 1, 7, and 11 are independent) are anticipated by U.S. Patent No. 8,340,975 to Rosenberger (Ex. 1003 or “Rosenberger”). Pet. 4. Petitioner treats claim 7 as representative of the independent claims and relies solely on an express anticipation theory. Pet. 24-48.

The Board should deny institution because Petitioner has failed to establish that Rosenberger anticipates, among other things, “*determining[/determine], [by the first computing device and] based at least in part on information received from the one or more additional computing devices, whether the first computing device should perform speech recognition on the spoken audio input,*” as recited in each of independent claims 1, 7, and 11. (emphases added). Furthermore, the ITC agrees that Rosenberger does not anticipate claim 7, specifically the “determine element” or any other element premised on the determination. *See* Ex. 2001, 49-50.

Petitioner cannot demonstrate a reasonable likelihood of succeeding for at least these reasons, as articulated below.

II. THE EVIDENCE AND ITC CONFIRM THAT ROSENBERGER DOES NOT DISCLOSE DETERMINING WHETHER TO PERFORM SPEECH RECOGNITION

A. Rosenberg's Control Device Performs Speech Recognition Before Information Is Received from Other Devices

Petitioner argues that Rosenberg discloses determining whether to perform speech recognition based at least in part on information received from the one or more additional computing devices because “Rosenberger’s ‘control device’ serving as the ‘coordinating controller’ (*first computing device*) is operable to determine, based at least in part on a ‘weighted signal’ (*information*) received from one or more additional ‘control devices’ (*one or more additional computing devices*) whether the ‘coordinating controller’ should perform speech recognition on *the spoken audio input*. ” Pet. 31 (emphases in Petition) (quoting Rosenberg, 10:34-38). Petitioner relies on various embodiments, including those illustrated in Figures 4 and 5A-5B, to demonstrate that the coordinating controller performs speech recognition based on a weighted signal received from another control device. Pet. 31-33. But none of these embodiments discloses a controller or coordinating controller that performs speech recognition based on a weighted signal—rather, the devices continuously perform speech recognition and determine whether to *further handle* speech commands based on the weighted signal. In sum, the “coordination”/ “determination” that Petitioner alludes to occurs after the

control device performs speech recognition on the received speech trigger / trigger phrase / triggering speech command.

As an initial matter, it is unclear which embodiment Petitioner is relying on to establish anticipation. For example, Petitioner cites Figure 4 but also cites several passages relating to Figures 5A-5B to explain how a control device “serving as the coordination controller” would operate. Pet. 31-33 (citing Rosenberger, FIGS. 5A, 5B, 12:28-13:10). But Figures 5A-5B and the cited passages relate to configurations in which a control device sends a weighted signal to an “independent” coordination controller that facilitates message transmission between controller devices, not a control device *serving as a coordination controller*. Rosenberger, 5:51-56 (“FIGS. 5A-5B are flowcharts illustrating preferred programming methods by which an independent coordinating controller that has received individual weighted signals from a plurality of control devices indicating the quality of a speech command each has just[]recognized decides which device will further process that speech command . . .”). Patent Owner submits that Petitioner’s anticipation analysis is fundamentally flawed because they have failed to establish that any single embodiment discloses the claims. *Microsoft Corp. v. Biscotti, Inc.*, 878 F.3d 1052, 1069 (Fed. Cir. 2017) (“[A]nticipation is not proven by ‘multiple, distinct teachings that the artisan might

somehow combine to achieve the claimed invention.” (citation omitted)); *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 851 F.3d 1270, 1274-75 (Fed. Cir. 2017) (the Board cannot “fill in missing limitations simply because a skilled artisan would immediately envision them”); *Net MoneyIN, Inc. v. VeriSign, Inc.*, 545 F.3d 1359, 1371 (Fed. Cir. 2008) (holding a system claim not invalid as anticipated where the prior art failed to disclose the limitations “arranged or combined in the same way as recited in the claim”).

To the extent Petitioner intended to rely on the embodiment illustrated in Figure 4, this embodiment fails to disclose performing speech recognition based on a weighted signal (alleged “information”) received from one or more additional control devices (alleged “one or more additional computing devices”) for at least the following reasons.

First, Rosenberger expressly discloses that each control device *continuously* performs speech recognition such that all computing devices in a building space recognize speech *upon receipt*—not based on a weighted signal:

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), . . .

the device . . . prompt[s] the user to say one of the speech
commands

Rosenberger, 8:20-32.

Second, Rosenberger expressly discloses that the control device broadcasts the weighted signal *after* the control device (alleged “computing device”) is “triggered,” i.e., after performing speech recognition on the received spoken audio input:

First, the microcontroller waits to be “**triggered[]**” (i.e., switched from a “listening” mode to an active mode). In the event the microcontroller *was triggered* hands-free *through its recognition* . . . , the device calculates a weighted signal . . . [and] *then sends its weighted signal* to each of the other interactive speech recognition control devices

Rosenberger, 11:40-48 (emphases added); *see also* Rosenberger, 5:45-50, 5:58-6:3 (“*control devices that have responded to the same speech command broadcast their weighted signals* and collectively decide which one will *further handle* such speech command”; “*control devices that have simultaneously recognized the same speech command send their weighted signals to each other* and collectively decide which one will *further handle* such speech command”; “control devices of the invention either coordinate amongst themselves or utilize a coordinating

controller to decide which control device will handle a given speech command

when more than one control device has been simultaneously triggered.”

(emphases added)). This sequence—in which **speech recognition** precedes

weighted signal (“information”) **transmission** or **receipt**—is illustrated in annotated

Figure 4 below:

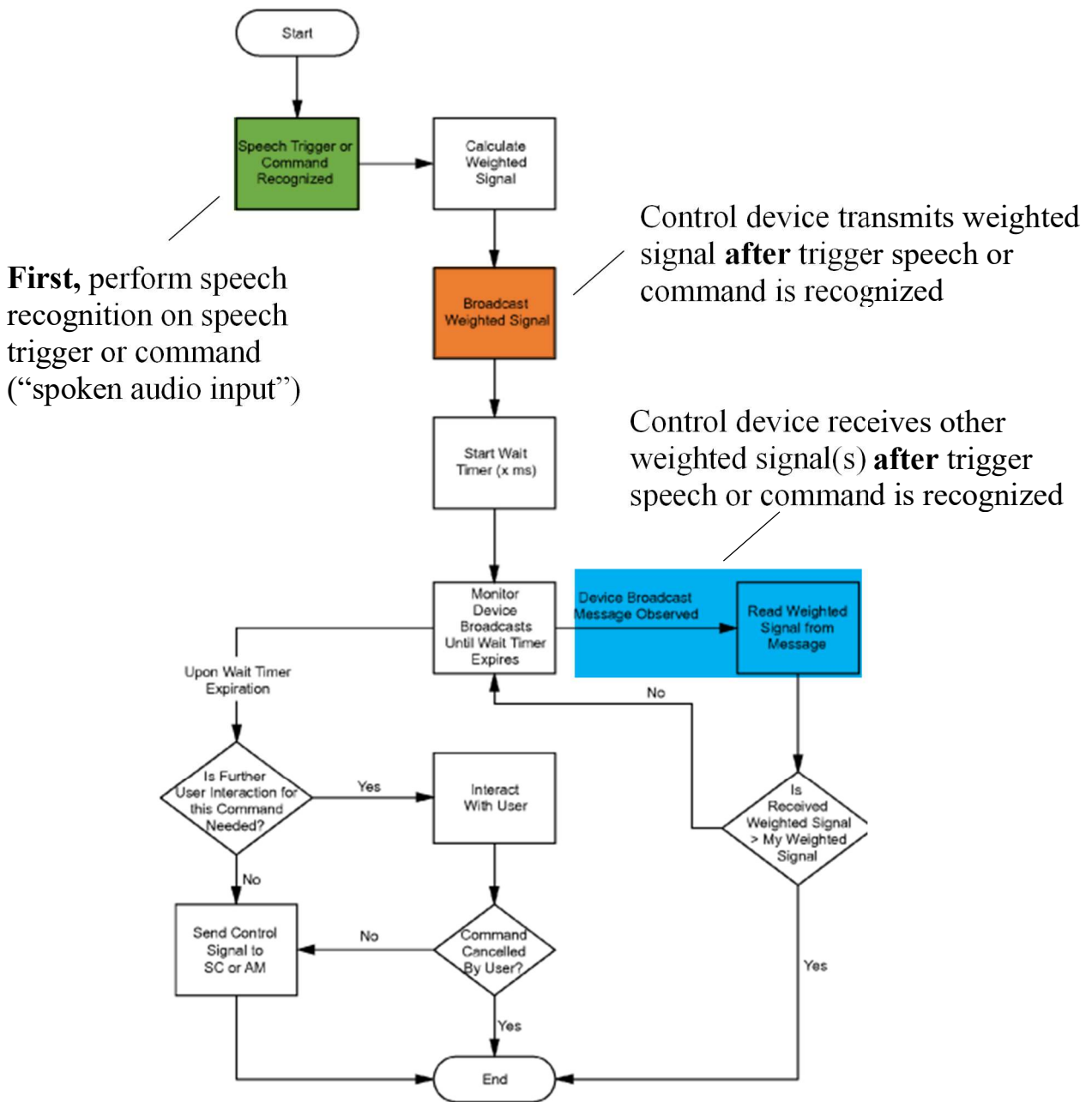


FIG. 4

Rosenberger, Fig. 4 (annotated).

Petitioner also acknowledges that the control device performs speech recognition on the received audio input *before* any weighted signals are calculated: “Rosenberger explains that all of its coordination methods involve ‘the calculation and production by the control device of a numeric weighted signal (WS) that quantifies the “quality” of a *just-recognized speech command*.’” Pet. 32 (emphasis added) (quoting Rosenberger, 10:34-41). And the claims further confirm that speech recognition is performed prior to sending any weighted signal, i.e., before the alleged determination is made:

(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*

Rosenberger, claims 1, 12, 19 (emphases added).

Thus, it cannot be true that the control device (even acting as a coordinating controller) performs speech recognition based on a received weighted signal. Accordingly, Rosenberger does not disclose *determining*, by a control device serving as a coordination controller (alleged “first computing device”) and based at

least in part on a weighted signal (alleged “information received from the one or more additional computing devices”), “whether the first computing device should perform speech recognition on the spoken audio input,” as required in the independent claims.

Petitioner's reliance on Figures 5A-5B is inapposite. As explained above, this embodiment relates to configurations in which a control device sends a weighted signal to an “independent” coordination controller that facilitates message transmission between controller devices, not a control device serving as a coordination controller. Rosenberger, 5:51-56.

Even so, Figure 5A plainly illustrates that speech recognition is performed on a received “[s]peech [t]rigger or [c]ommand” (alleged “spoken audio input”) *before* any weighted signals (alleged “information”) are transmitted or received by a control device or [coordinating controller](#). Rosenberger, 11:40-48, 12:28-35, FIG. 5A.

Coordinating controller receives/controller transmits weighted signal to coordination controller **after** trigger speech or command is recognized

First, perform speech recognition on speech trigger or command (“spoken audio input”)

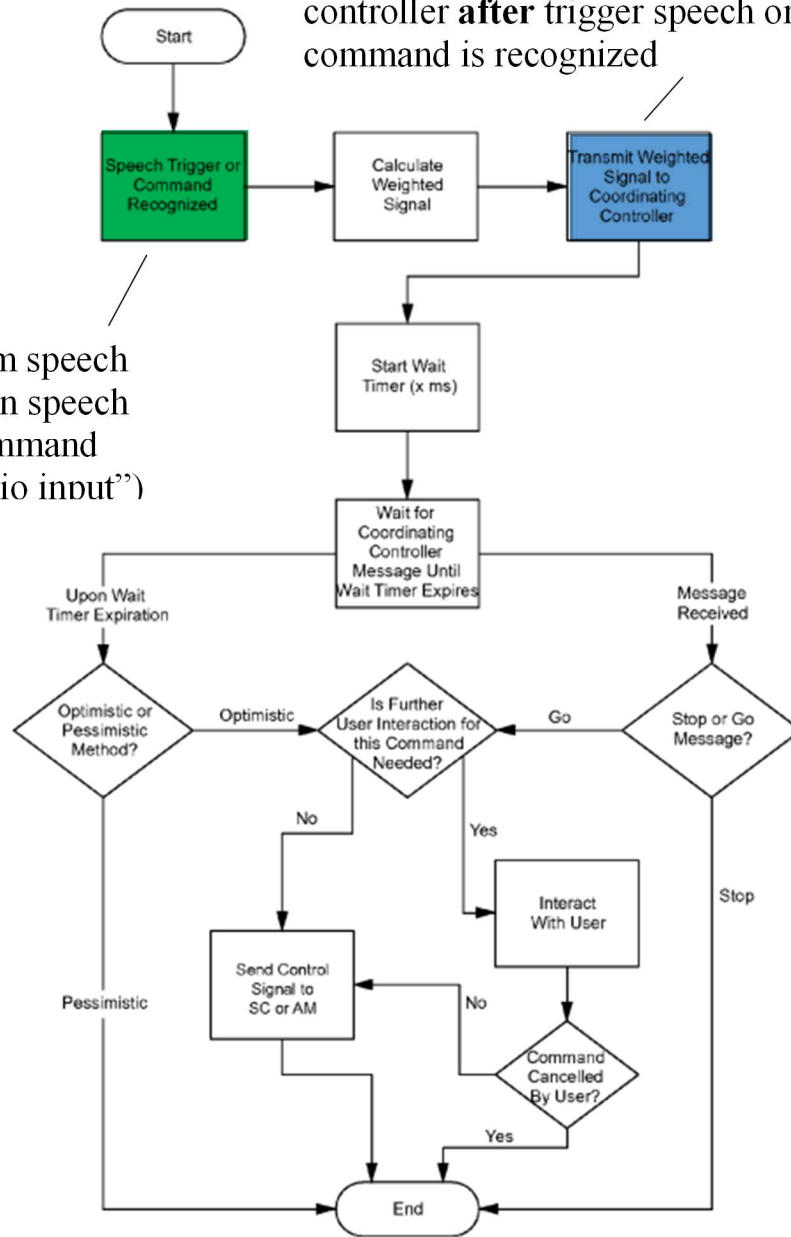


FIG. 5A

Rosenberger, FIG. 5A (annotated).

The coordination method illustrated in Figure 5B likewise confirms that the “coordination method[]” executed by the coordinating controller occurs *after* the control device transmits (and the coordinating controller receives) the weighted signals. Rosenberger, 10:34-38, 12:57-66. As illustrated below, the “[w]eighted [s]ignal [m]essage” is “ [o]bserved” and “[r]eceive[d]” by the coordinating controller *after* the control device is triggered and sends the weighted signal message. Rosenberger, 12:57-66, FIG. 5B.

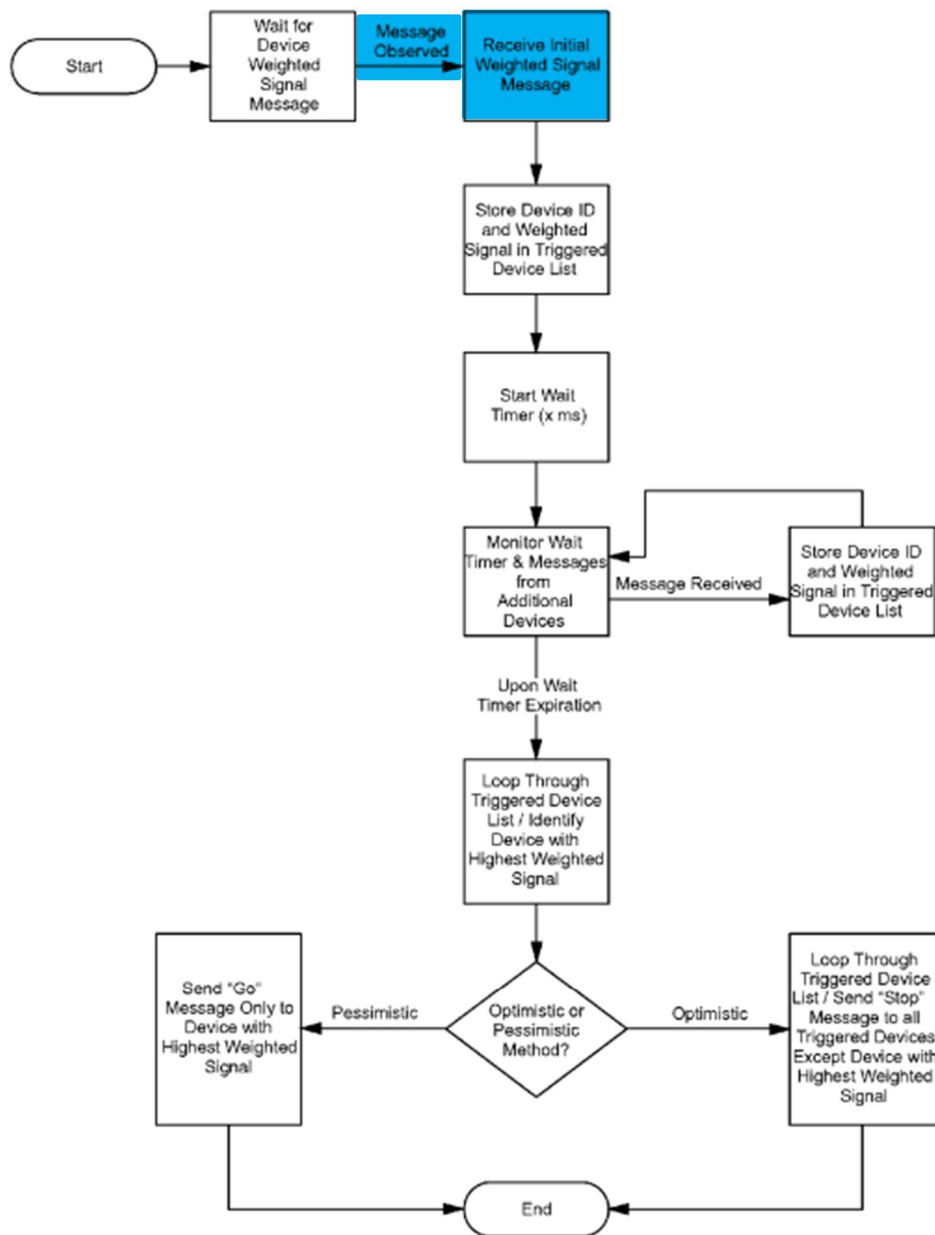


FIG. 5B

Rosenberger, FIG. 5B (annotated).

That the control device with the highest weighted signal “proceeds to handle the speech command by interacting with the user as needed and subsequently

transmitting the appropriate system control signals,” as argued by Petitioner (Pet. 33 (quoting Rosenberger, 11:55-12:3)), does not negate the fact that a control device already performed speech recognition prior to calculating the weighted signal. Thus, nothing in the description of the coordination method of Figures 5A-5B describes a coordinating controller performing coordination prior to performing speech recognition on the received trigger phrase or command, let alone determining “whether to perform speech recognition on the spoken audio input” “based at least in part on information received from the one or more additional computing devices.”

Accordingly, the embodiment of Figures 5A-5B does not anticipate “*determining[/determine], [by the first computing device and] based at least in part on information received from the one or more additional computing devices, whether the first computing device should perform speech recognition on the spoken audio input,*” as recited in each of independent claims 1, 7, and 11 (emphases added).

B. The ITC Agrees that Rosenberger Does Not Disclose Performing Speech Recognition Based on a Determination, as Claimed

In its Initial Determination on validity,¹ the ITC also found that Rosenberger's control devices continuously perform speech recognition without first making any determination to do so. Specifically, the ITC found that the devices, even those that do not handle a spoken command, "necessarily" perform "speech recognition on whatever it hears." Ex. 2001, 49-50 (citing Rosenberger, 11:63-64), 51 ("[T]he other devices continuously perform speech recognition on whatever they hear, even after they stop 'handl[ing] the speech command.'" (second alteration in original) (quoting Rosenberger, 11:62-67)).

Regarding Rosenberger's coordination method, the ITC found:

Even in the embodiment that uses a "trigger phrase," followed by a "speech command," the decision is over

¹ This discussion is limited to the ITC's particular findings regarding Rosenberger's alleged disclosure of determining whether to perform speech recognition. Patent Owner does not concede to the correctness of the ITC's construction of "perform[ing] speech recognition," particularly its finding that "hotword detection" necessarily includes speech recognition. Ex. 2001, 31. Patent Owner makes no representations in this POPR regarding the ITC's other findings as they relate to other Google patents.

which device “handles each discrete speech command,”
and *not over whether any particular device “perform[s]
speech recognition on the spoken audio input.”*

[Rosenberger] at 9:57-10:14 (describing the trigger
phrase “Hello Voice Control,” followed by the speech
command “Toggle All Shades”). . . . “[H]andl[ing]” the
speech command refers to the specific processing
involved in executing the command, including
“interacting with the user as needed and subsequently
transmitting the appropriate system control signals,” *not
speech recognition processing in general.* [Rosenberger]
at 11:67-12:2.

Ex. 2001, 50 (first and fourth alterations in original) (emphases added).

Accordingly, the ITC correctly determined that Rosenberger does not
disclose “the ‘determine’ element of claim 7” or the “entirety of the last group of
elements,” which are premised on the determining element. *Id.*

III. CONCLUSION

For all these reasons, Petitioner cannot meet its burden of establishing that at
least one claim is unpatentable as anticipated or obvious over the cited art, and the
Board should exercise its discretion to deny institution.

IPR2023-01288
U.S. Patent No. 9,632,748
Patent Owner's Preliminary Response

Respectfully submitted,

Dated: December 12, 2023

By: /Erika H. Arner/
Erika H. Arner, Reg. No. 57,540
Lead Counsel

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(d), I, Cory C. Bell, certify that this **PATENT OWNER'S PRELIMINARY RESPONSE** contains 2,039 words, excluding those portions identified in 37 C.F.R. § 42.24(a), as measured by the word-processing system used to prepare this paper.

Dated: December 12, 2023

By: /Cory C. Bell/
Cory C. Bell, Reg. No. 75,096

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. § 42.6(e), I certify that on December 12, 2023, a copy of this **PATENT OWNER'S PRELIMINARY RESPONSE** was served upon the below-listed counsel by electronic mail:

Michael P. Boyea
boyea@ls3ip.com

Jae Y. Pak
pak@ls3ip.com

Cole B. Richter
richter@ls3ip.com

LS3_team@ls3ip.com

Dated: December 12, 2023

By: /Lisa C. Hines/
Lisa C. Hines
Senior Litigation Legal Assistant
FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, LLP